Characterization of Putative Glycosyltransferases in Giant *Acanthamoeba Polyphaga* Mimivirus

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde (Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Anna Juliane Rommel

von

Winterthur ZH

Promotionskomitee

Prof. Dr. Thierry Hennet (Vorsitz)

Prof. Dr. Lubor Borsig

Prof. Dr. Urs Greber

Prof. Dr. Michela Tonetti

Zürich, 2016

TABLE OF CONTENTS

SUMMARY

The discovery of giant mimivirus in 2003 challenged the classical dogmas of virology: viruses were thought to be small, filterable agents that strongly depend on their host and contain a minimal number of genes for replication. Mimivirus with a particle size of 500 nm, however, is not only bigger in size than many parasitic or free-living microbes, mimivirus is also huge in genome size with 1.2 Mbp and 1'262 putative coding genes. Many of the mimivirus genes were never described in viruses before and were considered hallmark genes of cellular life. These genes feature tRNA synthases, enzymes involved in DNA repair, chaperones, and metabolic enzymes. The mimivirus genome also features a set of genes composing what appears to be a functional glycosylation machinery. Conventional viruses hijack the glycosylation machinery of their host and, with few examples, do not encode glycosyltransferases. There is only one example of another giant virus with a host independent glycosylation machinery known to date. Paramecium bursaria Chlorella virus encodes functional glycosyltransferases that are involved in the formation of virus-specific glycans for example on the major capsid protein. In mimivirus the host-independent synthesis for three nucleotide-diphospho sugars, namely UDP-N-acetylgulcosamine, UDP-rhamnose, and UDP-viosamine has been described. In 2011 a first glycosyltransferase of mimivirus was described. Mimivirus L230 encodes a bi-functional collagen-modifying enzyme hydroxylating lysine residues and glucosylating the latter. Nothing is known about the other eleven putative mimivirus glycosyltransferases. In this thesis, we investigated the functionality of six of these glycosyltransferases *in vitro*.

The putative glycosyltransferases were cloned and expressed in *E. coli* as His-fusion proteins. Protein expression and purification by affinity chromatography was possible for five glycosyltransferases, namely L193, R139, R363, R654, and R707. L373 was cloned and expressed in Spodoptera frugiperda Sf9 cells as FLAG-fusion protein. To gain an

idea about the possible functionality of our candidate genes, we analyzed the sequences using bioinformatics tools. Using the results from the sequence analysis together with the known monosaccharide composition of mimivirus and known O-glycan structures we tested the glycosyltransferase activity of the candidate proteins *in vitro*. Mimivirus R707 was identified as glycogenin-1 homolog that elongated glucose acceptor substrate with one or two glucose residues in a divalent metal ion dependent manner. The products of R707 catalysis featured α1,6 or β1,6 linkages as well as α1,4 linkages. A glycosyltransferase forming both alpha and beta glycosidic bonds has not been described before. Mimivirus L373 was identified as specific UDP-N-acetylglucosamine hydrolase with glycosyltransferase like characteristics. L373 activity was divalent metal ion dependent. We could rule out a number of activities for the other four putative glycosyltransferase candidates. Co-infection of amoeba with mimivirus and amylose or dextran induced a dose-dependent decrease of infection rate further underlining the importance of glycan glycan-binding protein interaction in mimivirus infection.

The results of this work gives further evidence for a mimivirus specific, host-independent glycosylation machinery and the necessity for mimivirus to invest energy in the production of virus specific glycans. The catalysis of R707 resulting in products with different configurations promote the peculiarity of mimivirus. Future in depth-study of mimivirus glycosyltransferase catalysis might give valuable insights on mechanisms of glycosylation in general.

ZUSAMMENFASSUNG

Die Entdeckung des Riesenvirus Mimivirus 2003 stellte klassische Dogmen der Virologie in Frage. Viren wurden bis dato als kleine, filtrierbare Stoffe eingestuft, die stark von ihrem Wirt abhängig sind und eine minimale Anzahl von Genen mitbringen. Mimiviruspartikel haben einen Durchmesser von 500 nm und sind damit grösser als viele parasitäre oder frei lebende Mikroben. Ausserdem ist das Genom von Mimivirus mit einer Grösse von 1.2 MBp und 1‘262 mutmasslichen Genen riesig. Viele dieser Gene wurden noch nie in Viren beschrieben und vorgängig als exklusive Gene des zellulären Lebens gesehen. Einige Beispiele sind tRNA-Synthasen, DNS-Reparaturenzyme, Chaperone und Stoffwechselenzyme. Ausserdem enthält das Genom von Mimivirus mehrere Gene, die eine scheinbar voll funktionsfähige Glykosylierungsmaschinerie bilden. Konventionelle Viren nutzen die Glykosylierungsmaschinerie ihres Wirts. Die einzige heute bekannte Ausnahme bildet Paramecium bursaria Chlorellavirus. Dieses Riesenvirus kodiert funktionelle Glykosyltransferasen, welche für die Synthese von virusspezifischen Glykanen zum Beispiel auf dem viralen Kapsidprotein verantwortlich sind. Die Synthese von mehreren Nukleotidzuckern, nämlich UDP-N-acetylglucosamine, UDP-Rhamnose und UDP-Viosamine, in Mimivirus ist bereits beschrieben. 2011 wurde eine erste, funktionelle Glykosyltransferase in Mimivirus charakterisiert. Mimivirus L230 kodiert ein bifunktionales, Kollagen modifizierendes Enzym, welches Lysine hydroxyliert und diese anschliessend glukosyliert. Über die weiteren elf mutmasslichen Glykosyltransferasen in Mimivirus ist bis heute nichts bekannt. In Rahmen dieser Arbeit wurden die Funktionalität von sechs der elf mutmasslichen Glykosyltransferasen und deren mögliche Rolle in der Infektion von Amöben untersucht.

Die sechs Glykosyltransferasen wurden als rekombinante Proteine in *E. coli* exprimiert. Die Expression und Aufreinigung der Proteine war in fünf von sechs Fällen erfolgreich,

nämlich im Fall von L193, R139, R363, R654 und R707. Der sechste Kandidat, L373, wurde in Spodoptera frugiperda Sf9 Zellen exprimiert. Um eine Idee von der Funktionalität der Glykosyltransferasekandidaten zu bekommen, analysierten wir die Proteinsequenzen mit Hilfe von bioinformatischen Programmen. Die so gewonnenen Resultate in Kombinaten mit schon bekannten Daten über die Monosaccharid-komposition und O-Glykanstrukturen von Mimivirus nutzten wir für gezielte Tests der Glykosyltransferaseaktivität *in vitro*. Mimivirus R707 wurde als Glykogenin-1 Homolog identifiziert, welches Glukoseakzeptorsubstrate mit einem oder zwei Glukoseeinheiten verlängerte. Die Produkte der R707 Katalyse hatten sowohl β1,6 als auch α1,6 und α1,4 Bindungen. Die Katalyse war Metallionen abhängig. Eine Glykosyltransferase, die Alpha- und Betabindungen formt, wurde bis anhin noch nicht beschrieben. Mimivirus L373 wurde als spezifische UDP-N-acetylglukosaminylhydrolase mit Glykosyl-transferaseeigenschaften identifiziert. Die Aktivität von L373 war Metallionen abhängig. Für die anderen mutmasslichen Glykosyltransferasen konnten wir mehrere mögliche Aktivitäten ausschliessen. Die parallele Infektion von Amöben mit Mimivirus und Amylose oder Dextran führte zu einer konzentrationsabhängigen Reduktion der Infektionsrate. Dieses Resultat unterstreicht die Wichtigkeit der Interaktion von Glykanen und Glykan bindenden Proteinen während der Mimivirusinfektion.

Im Rahmen dieser Arbeit wurden weitere Beweise dafür erbracht, dass Mimivirus eine spezifische, wirtsunabhängige Glykosylierungsmaschinerie besitzt und es für Mimivirus lohnenswert ist, Energie in eine eigene Glykosylierungsmaschinerie zu investieren. R707 Katalyse resultierte in Produkten mit verschiedenen Konfigurationen und unterstreicht damit Eigentümlichkeit von Mimivirus. Zukünftige Untersuchungen der Katalyse-mechanismen von mimiviralen Glykosyltransferasen könnten wertvolle Erkenntnisse über Mechanismen von Glykosyltransferasen im Allgemeinen geben.

LIST OF ABBREVIATIONS

Aa	amino acid
Amp	ampicillin
AP	*Acanthamoeba polyphaga*
Asn	asparagine
BLAST	basic local alignment search tool
BSA	bovine serum albumin
CAZY	carbohydrate active enzyme database
CID-MS	collision-induced dissociation mass spectrometry
CRISPR	clustered regularly interspaced short palindromic repeats
CroV	*Cafeteria rhoenbergensis* virus
DMSO	dimethylsulfoxide
DQF-COSY	double quantum filtered correlated spectroscopy
DTT	dithiothreitol
ER	endoplasmic reticulum
ESI-MS	electrospray ionization mass spectrometry
Fuc	fucose
Gal	galactose
GC-MS	gas chromatography-mass spectrometry
Glc	glucose
GlcN	glucosamine
GlcNAc	N-acetylglucosamine
GlcNAc	N-acetylglucosamine
GYG1	glycogenin-1
HA	hemagglutinin
HCD-MS	higher-energy collisional dissociation mass spectrometry
HexNAc	N-acetylhexosamine
HIV	human immunodeficiency virus

HMBC	heteronuclear multiple bind correlation
HPAEC-PAD	high-performance anion-exchange chromatography with pulsed amperometric detection
HSQC	heteronuclear single quantum coherence
Hyl	hydroxylysine
Kan	kanamycin
LPS	lipopolysaccharide
MALDI-MS	matrix-assisted laser desorption/ionization mass spectrometry
Man	mannose
NCLDV	nucleocytoplasmic large DNA virus
Neo	neomycin
NMR	nuclear magnetic resonance
ORF	open reading frame
PBCV	*Paramecium bursaria* chlorella virus
PBS	phosphate buffered saline
Phyre^2	protein homology/analogy recognition engine v 2.0
pNP	4-nitrophenyl
Rha	rhamnose
ROESY	rotating frame Overhauser enhancement spectroscopy
SEM	scanning electron microscopy
TEM	transmission electron microscopy
TOCSY	two-dimensional nuclear magnetic resonance spectroscopy
UDP	uridine diphosphate
VF	virus factory
Vio	viosamine
Xyl	xylose

INTRODUCTION

In the past decade many members of a new, exciting class of viruses were discovered. These so-called giant viruses represent a curiosity in nature and their discovery challenges the dogma that viruses are small, filterable agents which fully depend on their hosts. Giant viruses harbor a variety of genes never seen in viruses before. Giant viruses can replicate with only small dependence on their hosts. *Acanthamoeba polyphaga* mimivirus was the first member of giant viruses to be discovered and has been their model organism ever since. Amongst its many unusual genes mimivirus encodes its own glycosylation machinery comprised of several nucleotide sugar synthesizing and modifying enzymes, and eleven genes that share structural similarity with prokaryotic and eukaryotic glycosyltransferases. Comprehension of the glycosylation machinery of mimivirus will contribute to understanding the role of glycosylation in giant virus infection and replication.

ACANTHAMOEBA POLYPHAGA MIMIVIRUS

DISCOVERY

Acanthamoeba polyphaga mimivirus was originally discovered after a pneumonia follow-up study in 1992. Mimivirus was isolated from water from a cooling-tower in England and thought to be a small Gram-positive bacterium named Bradfordcoccus (1). After "Bradfordcoccus" was studied within its natural host and after failed amplification of a 16S rRNA, the thought to be bacterium revealed virus characteristics. In electron microscopy studies mimivirus presents an icosahedral capsid with a diameter of 400 nm surrounded by 140 nm long protein fibers (Figure 1C). Oligosaccharide coating of the fibers is dense and likely explaining the positive Gram-staining of mimivirus particles. Bradfordcoccus was renamed mimivirus, short for microbe mimicking virus (2).

PARTICLE STRUCTURE

The icosahedral capsid of mimivirus is covered with heavily glycosylated surface fibers. These fibers commonly share the same length and are attached to anchor proteins. An approximately 25 kDa protein head group is attached to each fiber (Figure 1A/B/D). Surface fibers are produced independently of the virus particles and attached to the naked virion in a later step. Between the base of the surface fibers and the capsid protein a layer of connecting material is found. This anchor layer is seen as more or less regular, closely packed array of protein units of similar size. The anchor proteins protrude 3-4 nm above the capsid surface and reside in the depressions of the capsid mesh (Figure 1A/B/F). There are more than 3'000 fiber anchor proteins found on the surface of one mimivirus particle and each anchor protein allows the binding of multiple surface fibers. The major capsid protein of mimivirus forms an open network with a honeycomb structure of alternating hexagonal rings and depressions (Figure 1E). On each virus particle one five-fold vertex is occupied by a so-called "stargate". The stargate allows delivery of mimivirus DNA into the host cytoplasm. The stargate arms are not coated with anchoring proteins. Each of the five arms is 250 nm long, 50 nm wide and protrudes above the capsid by 20 nm (Figure 1A/B/F). No capsid protein lies beneath the arms of the stargate. Only little substructure is found on the surface of the stargate arms and the arms are very resistant to proteases. The stargate is closed by a protein plug. Upon opening some of the arms are shed and the five capsid triangles fold back allowing the interior sac to emerge. The interior sac is composed of host derived lipid bilayer membrane and contains the viral DNA. With a diameter of 340 nm the interior sac has a spherical volume of 2.1×10^7 nm^3. Packing of the 1.2 Mbp DNA in this volume would result in a packing density of 0.06 nm^3/ bp, which is relatively low. This allows the accommodation of bound and unbound protein in the interior sac. The DNA is not twisted around itself or tightly bundled but present as a single strand with loops. The DNA is highly associated with proteins of various sizes and shapes (Figure 1A/B/G).

Mimivirus DNA, DNA associated proteins, and the membrane sac occupy 65% of the capsid volume. The inside surface of the capsid is coated with a near continuous layer of proteins or a lipid membrane associated with proteins. The remaining space between the capsid and the interior sac is filled with additional material. After the disruption of virions fibers emerge that cannot be assigned as being DNA or RNA. These fibers are usually organized into cables, bundles, or coils. A single fiber has a diameter of 1 nm and a prominent 7 nm periodicity of dark and light which indicates a helical structure. The fibers form many stranded ribbons with transverse 7 nm periodic bands. The ribbons are flexible and form coils. The coils are not branched and not associated with other proteins. The fiber coils are hundreds of nm long and their origin is not known (Figure 1A/B/H) (3).

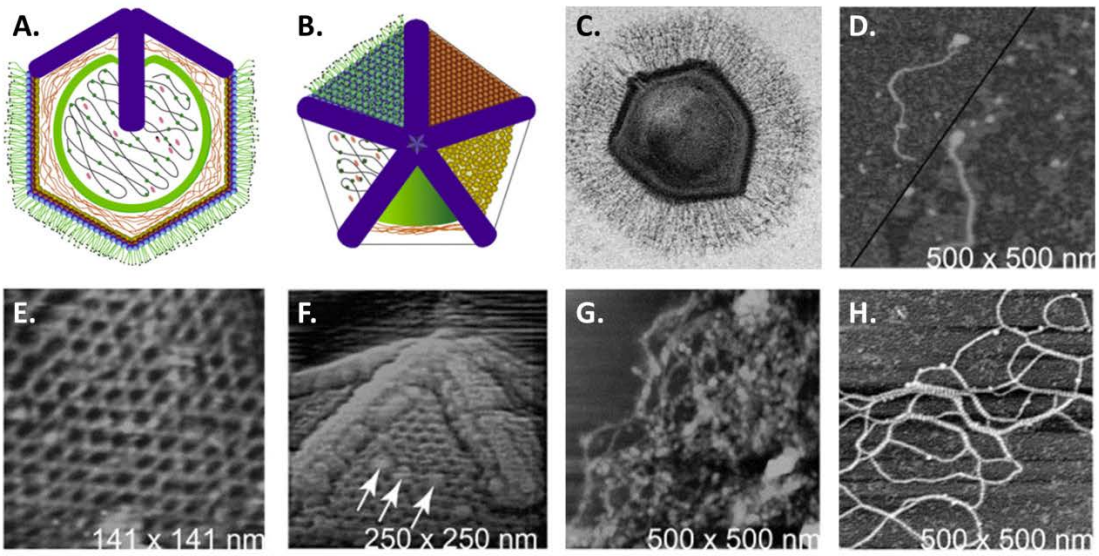


Figure 1 Mimivirus structures (adapted from (3-5)): A. and **B.** Schematic cross-section of mimivirus perpendicular to the fivefold axis of the particle and schematic top view of mimivirus with its prominent stargate (purple). From exterior to interior (**A**) or clockwise (**B**): head proteins (black) of surface fibers (green) attached to anchor protein (blue spheres) covering the capsid lattice (red spheres). Underneath the capsid an additional layer of protein (or a lipid membrane

studded with proteins) is found (yellow spheres). Between the surface of the capsid and the lipid bilayer (light green) fibers of intermediate length (orange) are seen. Inside the membrane are the genomic DNA associated with proteins (green) and other proteins not directly bound to the DNA (pink). **C.** Isolated mimivirus particle. Scale bar 200 nm. **D.-H.** Atomic force microscopy pictures of various mimivirus structures. **D.** Single mimivirus surface fibers with their head proteins and oligosaccharide coat. **E.** Mimivirus capsid protein forming a honeycomb network. **F.** Mimivirus stargate rising above the capsid protein mesh. Fiber anchor proteins marked with white arrows. **G.** Mimivirus DNA that is heavily associated with various proteins. **H.** Irregular networks and ribbons of fibers.

GENOME ORGANIZATION

The double-stranded DNA genome of mimivirus has a length of 1.2 Mbp (GenBank Accession number: NC_014649.1) and contains 1'262 putative open reading frames (nos) of which 979 are expected to be protein-coding genes, 6 tRNAs, and 33 ncRNAs (6). The linear genome contains two inverted repeats of 900 nucleotides near the termini. The DNA consists of 72% A+T. The gene distribution amongst the two genome strands is similar (450 on "R" and 461 on "L" strand, short for right and left strand) (7). Mimivirus contains many ORFs previously identified in other members of nucleocytoplasmic large DNA viruses (NCLDVs), yet a lot of genes were never identified in viruses before. It is therefore postulated that mimivirus represents a first member of Mimiviridae, a new family within the NCLDV (8). The unique genes of mimivirus encode proteins involved in protein translation (e.g. tRNA-like genes, aminoacyl tRNA synthetases, translation initiation factors, and peptide chain release factor), DNA repair (e.g. formamido-pyrimidine-DNA glycosylase, UV-damage endonuclease, enzymes involved in DNA mismatch repair), topoisomerases, chaperones (e.g. heat shock 70-kD, peptidylprolyl isomerase, and ion domain protease), and new enzymatic pathways (e.g. genes involved in glutamine metabolism, glycosyltransferases, and lipid-manipulating enzymes) (Table 1) (7). Some of these novel virus genes were classified as hall-mark genes for cellular

organisms. This makes it especially interesting to study the behavior of mimivirus in its host *Acanthamoeba polyphaga* (AP).

Table 1 Selection of new features identified in mimivirus genome (adapted from (7))

ORF no.	Definition/putative function	Comment
R663	Arginyl-tRNA synthetase	Translation
L124	Tyrosyl-tRNA synthetase	Translation
L164	Cysteinyl-tRNA synthetase	Translation
R639	Methyonyl tRNA synthetase	Translation
R726	Peptide chain release factor eRF1	Translation
R624	GTP-binding elongation factor eF-Tu	Translation
R464	Translation initiation factor SUI1	Translation
L359	DNA mismatch repair ATPase MutS	DNA repair
R693	Methylated-DNA-protein-cysteine methyltransferase	DNA repair
R406	Alkylated DNA repair	DNA repair
L687	Endonuclease for the repair of UV-irradiated DNA	DNA repair
L254 L393	Heat shock 70-kDa	Chaperonin
L605	Peptidylprolyl isomerase	Chaperonin
R418	NDK synthesis of nucleoside triphosphates	Metabolism
R475	Asparagine synthase (glutamine hydrolyzing)	Metabolism
R565	Glutamine synthetase (Glutamate-amonia ligase)	Metabolism
R689	N-acetylglucosamine-1-phosphate, uridyltransferase	Polysaccharide synthesis
L136	Sugar transaminase, dTDP-4-amino-4,6-dideoxyglucose biosynthesis	Polysaccharide synthesis
L780	dTDP-4-dehydrorhamnose reductase	Polysaccharide synthesis

REPLICATION CYCLE

The necessary particle diameter for phagocytosis in *Acanthamoeba* is 0.6 μm or larger (9). Mimivirus infection is initiated by phagocytosis, mediated by actin protrusions resulting in phagosome-enclosed virions (Figure 2A) (4, 10). After lysosomal fusion the stargate opens allowing viral capsid to fold back (Figure 2B). Stargate opening enables membrane fusion between the phagosome and the inner membrane sac of mimivirus generating a large portal through which mimivirus genome content is released (10, 11). The penetration of the cytoplasm is followed by the disappearance of mimivirus for about two hours. After this eclipse phase the virus factory (VF), a structure within the cytoplasm of AP where mimivirus replication takes place, begins to occur (Figure 3). The VF demands for highly ordered trafficking of multiple viral proteins, that are expressed in the host cytoplasm (11). Within six hours the VF takes up a large volume of the amoebal cytoplasm (Figure 3 and Figure 2D). AP mitochondria are spatially associated with the VF, reflecting the need of mimivirus for energy (11). Six to nine hours post-infection the late phase of mimivirus replication takes place accompanied by icosahedral procapsid undergoing DNA packaging (Figure 2C). It is postulated that stargate formation occurs at an early stage of viral assembly. Biogenesis of the internal membrane sac is initiated by fusion of multiple small vesicles derived from host endoplasmic reticulum (ER) finally forming large single-layered membranes (Figure 2E).

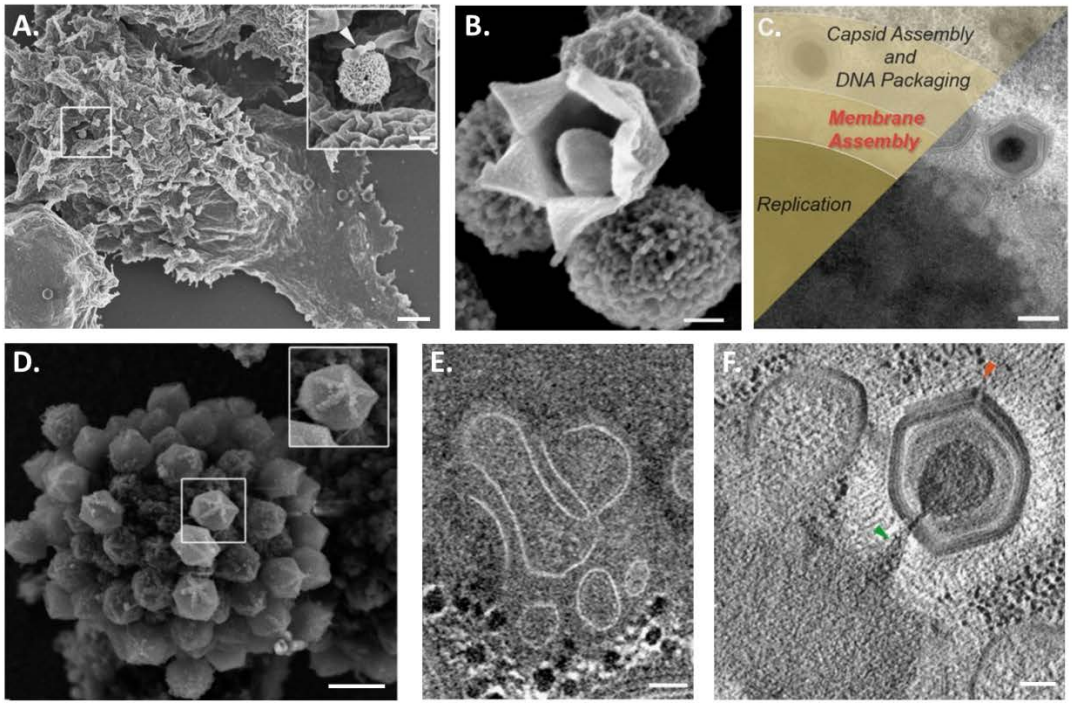


Figure 2 Important steps of mimivirus life cycle (adapted from (11)). A. SEM of mimivirus entry into amoeba via phagocytosis mediated by actin protrusions (arrowhead), scale bar 2 μm, zoom in 400 nm **B.** Extracellular mimivirus particle revealing the structure of the stargate five-fold opening and the interior membrane sac, scale bar 100 nm **C.** TEM of cytoplasmic virus factory depicting three distinct assembly zones. Scale bar 500 nm **D.** SEM of viral factory isolated 8 h post infection. Scale bar 500 nm **E.** Tomogram slice of multi-vesicular membrane structures generated at the periphery of mimivirus factory serving as source for the viral internal membrane layer. Scale bar 50 nm **F.** Tomogram slice of capsid undergoing DNA packaging with the packaging portal (green arrowhead) on the distal site of the stargate (red arrowhead). Scale bar 100 nm

DNA packaging in pre-assembled capsids takes place at the distal site of the stargate, meaning DNA exit and packaging proceed through different portals, a unique feature amongst genome translocation processes and one of the features separating mimivirus from other NCLDV (10) (Figure 2F).

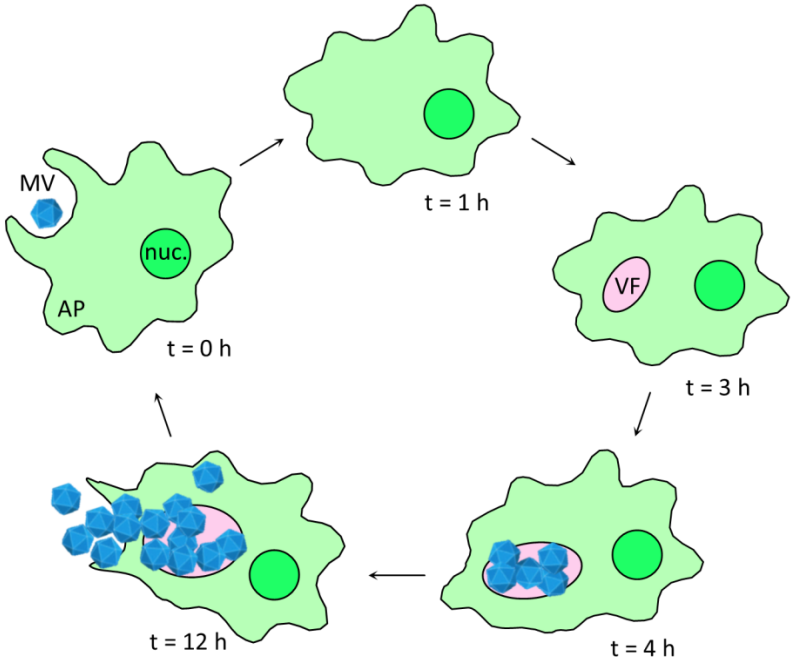


Figure 3 Mimivirus replication cycle in *Acanthamoeba polyphaga* (AP). After phagocytosis by AP, mimivirus (MV) disappears during an eclipse phase. After three hours the virus factory (VF) starts to appear. The VF expands within the following nine hours taking up more space in the host cytoplasm as MV is replicating. Finally AP bursts open releasing new MV particles. Nuc = AP nucleus.

NUCLEOCYTOPLASMIC LARGE DNA VIRUSES

As mentioned above, after its discovery and sequencing, mimivirus was classified as first member of the family of Mimiviridae within the monophyletic group of NCLDVs. Some other Mimiviridae family members discovered in the following years are Megavirus chilensis (12), A. castellanii Mamavirus (13), Moumouvirus Monve (14), and Cafeteria rhoenbergensis virus (15) (see below). NCLDVs unite 7 families of viruses: Poxviridae, Iridoviridae, Ascoviridae, Asfarviridae, Marseilleviridae, Phycodnaviridae, and Mimiviridae, plus the recently discovered, unclassified Pandoraviruses and

Pithoviruses (Figure 4) (16). An official taxonomic rank was proposed to be assigned to NCLDVs: the order of "Megavirales" (17).

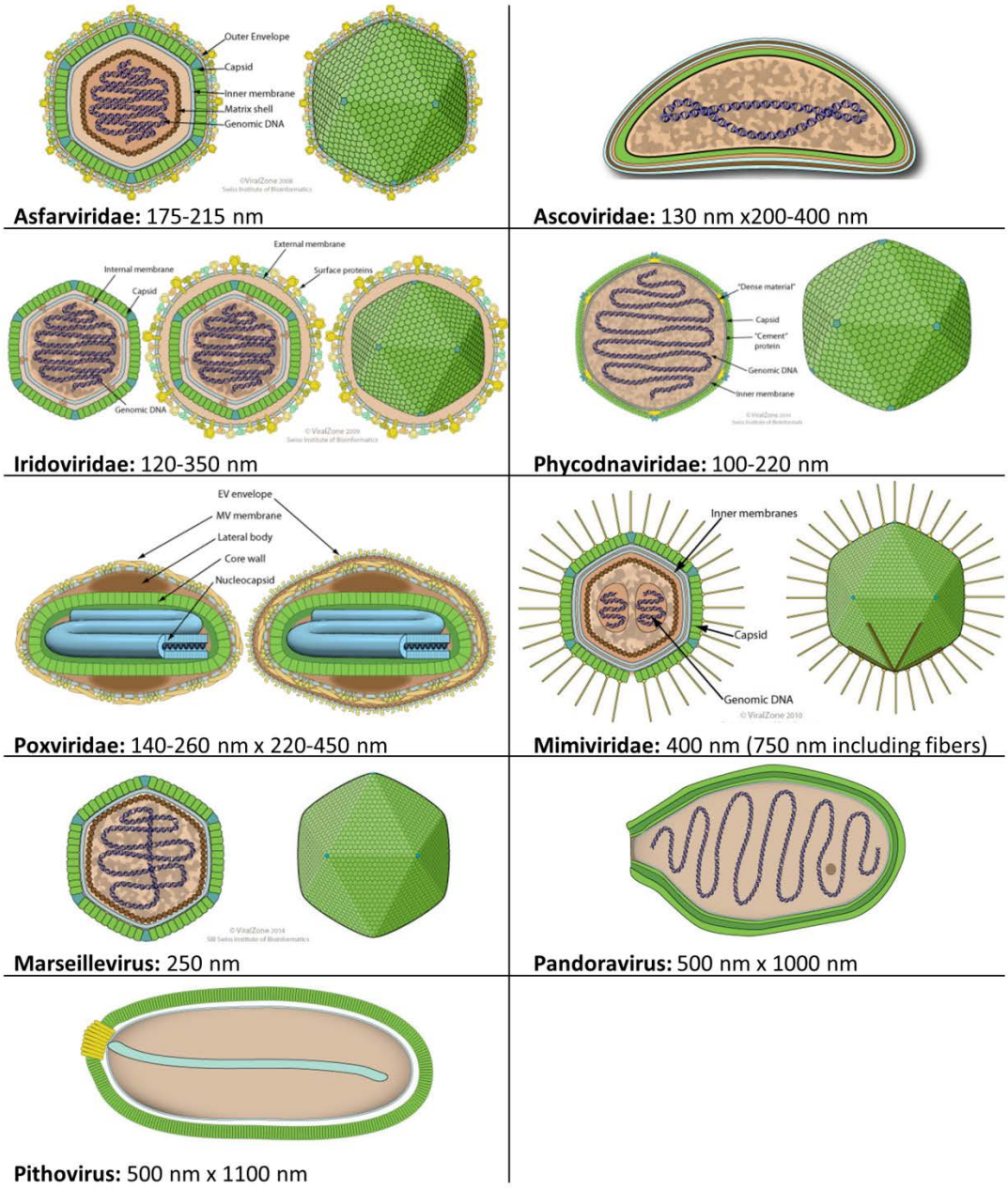


Figure 4 Schematic representation of major members of NCLDV families known to date (adapted from (18)). Green: Capsid protein.

NCLDVs feature a double-stranded DNA genome and replicate either exclusively in the cytoplasm or start their replication in their host's nucleus and finish their life cycle in the cytoplasm (Table 2). NCLDVs infect a wide range of hosts (Table 2). The common ancestor of the NCLDVs contains a conserved set of 47 core genes. In 2009, an analysis of 45 annotated proteomes of NCLDVs partitioned 9'261 of the 11'468 viral proteins in 1'445 clusters of probable orthologs. Amongst these 1'445 clusters of orthologs 1'268 clusters are family specific. The other 177 clusters of orthologs contain proteins from two or more NCLDV families. These widespread clusters of orthologs consist of genes encoding key functions of viral replication and morphogenesis (19). Most NCLDVs have capsid proteins composed of a jelly-roll fold, which is composed of a wedge-shaped, eight-stranded, anti-parallel β-barrel. NCLDVs have an internal membrane sac containing the viral DNA (20).

Table 2 Features of the NCLDV families (adapted from (21))

Virus family	Host	Genome size range (kb)	Genome structure	Replication site
Asfarviridae	Mammals	170	Linear	Cytoplasm
Ascoviridae	Insects	150-190	Circular	Nucleus and cytoplasm
Iridoviridae	Insects, cold-blooded vertebrates	100-220	Linear	Nucleus and cytoplasm
Phycodnaviridae	Green-algae, algal symbionts of paramecia and hydras	150-400	Linear/ circular	Nucleus and cytoplasm
Poxviridae	Insects, reptiles, birds, mammals	130-380	Linear	Cytoplasm
Mimiviridae	*Acanthamoeba*, probably also algae	700-1'200	Linear	Cytoplasm
Marseilleviridae	*Acanthamoeba*, probably also algae	370	Circular	Nucleus and cytoplasm

Putative family including Pandoravirus (22)	*Acanthamoeba*	1’900-2’500	Linear	Nucleus and cytoplasm
Pithovirus sibericum (23)	*Acanthamoeba*	610	Circular	Cytoplasm

ASFARVIRIDAE

Asfarvirus is an enveloped virus with an icosahedral capsid and a nucleoprotein complex (Figure 4). The only member of Asfarviridae causes African swine fever in domestic pigs and their relatives. The main symptom is hemorrhagic fever with a mortality rate of up to 100%. There is no species known to infect humans (24).

ASCOVIRIDAE

Ascoviridae are enveloped viruses with an internal core consisting of an intra-capsid lipid membrane and an inner particle (Figure 4). They are bacilloform, ovoid or allantoid and infect mainly invertebrates. Ascoviridae evolved from Iridoviridae or from Phycodnaviridae. Ascoviruses show some unusual features: cell infection induces the development of apoptotic bodies that are then used for viral assembly (25).

IRIDOVIRIDAE

Iridoviridae consist of an outer capsid, an intermediate lipid membrane and a central core containing DNA-complexes. The virions have an icosahedral form (Figure 4). Iridoviridae infect a broad-range of invertebrates and cold-blooded vertebrates. A prominent member is ranavirus infecting amphibians and being the second most widespread infectious disease of amphibians (26).

PHYCODNAVIRIDAE

Phycodnaviridae contain an inner lipid membrane and have an icosahedral capsid structure (Figure 4). They infect marine and freshwater algae, having an impact on geochemical cycling and weather patterns (27). An interesting member is Paramecium bursaria chlorella virus (PBCV), the pioneer amongst viruses with genomes larger than 300 kb (28), which also encodes its own glycosylation machinery (see below) (29). Phycodnaviridae evolved from Iridoviridae (27).

POXVIRIDAE

Poxviridae are enveloped viruses which have a distinct, brick-shaped form originating from the outer membrane layer formed by the host's ER (Figure 4). Poxviruses produce two types of infectious particles, mature virions and extracellular virions (30). Famous representatives are vaccinia virus and the small pox virus. The world's first successful antiviral vaccine was developed against smallpox virus by using vaccinia virus, a close relative to poxvirus from horse and cow, as immunogen (31).

MARSEILLEVIRIDAE

Marseillevirus, the first member of this family, was discovered in 2007. Marseilleviridae have an icosahedral capsid and encode approximately 500 proteins (Figure 4). The family includes minimum 17 members, of which 14 are unassigned. The most prominent members are AP Marseillevirus and *Acanthamoeba castellanii* Lausannevirus (32-34). Antibodies against marseilleviruses have been found in the blood of asymptomatic humans, raising the question, whether marseilleviruses might be pathogenic in humans (34).

PANDORAVIRUSES

Pandoraviruses are the biggest viruses discovered to this point. Their genomes range from 2.5-2.8 Mbp. The particles are amphora-shaped (Figure 4). The first two members

discovered were Pandoravirus salinus, found in Chile and Pandoravirus dulcis discovered in a fresh-water pond near Melbourne (22).

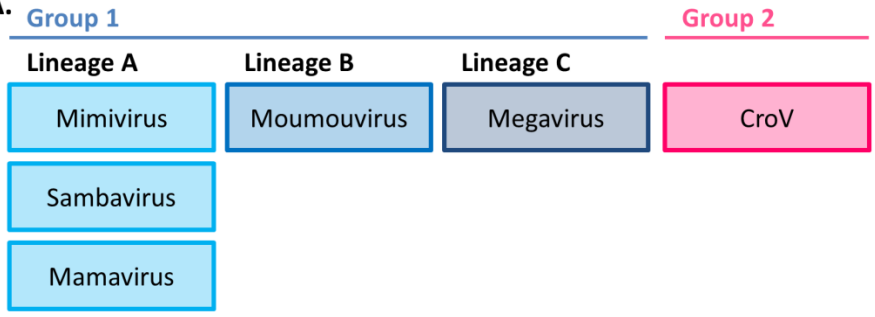
PITHOVIRUS SIBERICUM

Pithovirus sibericum was isolated from more than 30'000 year old Siberian permafrost. The virus shows Pandoravirus-like morphology and size, but its genome is more similar to Iridoviruses and marseilleviruses and with 600 kb rather small in comparison to the size of the virion (Figure 4) (23).

MIMIVIRIDAE

Mimiviridae are classified in two groups, whereof group 1 is divided in three lineages (A, B, C) (Figure 5). Lineage A contains mimivirus itself plus Mamavirus, Sambavirus, and others like Terra2 (35). Mamavirus was originally isolated from water from a cooling tower in Paris in 2008. The Mamavirus genome is slightly bigger than the genome of mimivirus. The genomes are highly similar but show some divergences towards the termini. The Mamavirus genome also contains some unique genes like a small regulatory polyA polymerase subunit that is similar to poxvirus genes (13). Sambavirus together with its virophage was discovered in Rio Negro River in Brazil in 2014. The genome of Sambavirus is about 50 kbp larger than the mimivirus genome (36). Terra2 was isolated from soil from Marseille (37). Moumouvirus is a member of lineage B (35). Moumouvirus was isolated from a cooling tower in southeastern France (14). Several viruses form lineage C: Megavirus chilensis, Courdo7, Courdo11, Terra1, Montpellier, and others (35). Megavirus was isolated from ocean water off coast of Chile, but it is capable of infecting fresh water *Acanthamoeba*. The Megavirus genome is slightly bigger than the mimivirus genome and it encodes three additional aminoacyl-tRNA genes (12).Terra1, like Tera2, was isolated from soil sample collected in Marseille in 2009 (37). More and more family members are being discovered over time (38-41).

A.



B.

Virus	Genome size (bp)	Capsid diameter (fiber length) (nm)	Putative coding genes	GenBank Accession #
Megavirus	1'259'197	500-550 (75)	1'120	NC_016072
Sambavirus	1'213'607	500-600 (90-120)	938	KF959826
Mamavirus	1'191'693	500-600 (100-200)	1'023	JF801956
Mimivirus	1'181'404	500-600 (100-200)	979	NC_014649
Moumouvirus	1'021'348	400-450 (100)	930	NC_020104
CroV	617'453	300	544	GU244497

Figure 5 Groups and lineages within the family of *Mimiviridae* (A) compared in genome and particle size (B).

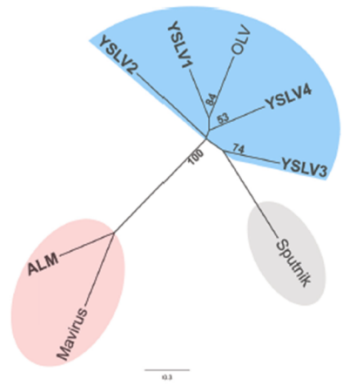
Group 2 within the family of Mimiviridae is represented by a more distant relative of mimivirus, Cafeteria rhoenbergensis virus (CroV). Less than one third of the CroV genes have homologs in mimivirus. CroV was isolated in 2010 from a coastal water sample collected off Texas in the early 1990ies (15).

VIROPHAGES OF GIANT VIRUSES

When Mamavirus was discovered co-infection of the giant virus with a smaller, icosahedral virus was observed. The virophage sputnik has an 18-kbp double-stranded, circular genome encoding 21 proteins (42). Sputnik cannot multiply in AP but rather hijacks the virus factories formed by Mamavirus (43). Other virophages of giant viruses

were discovered such as Mavirus which is associated with CroV. Mavirus is a dsDNA virus with a 19 kb genome encoding 20 proteins (44). Comparative genomic and phylogenetic analyses lead to the assumption, that at least three major virophage lineages are associated with NCLDVs (Figure 6) (45).

A.



B.

Virophage	Host Virus	Host Eukaryote	Genome Size [kbp]	Genome No. of ORF
Sputnik	Mimivirus	*A. polyphaga*	18.3	21
Mavirus	CroV	*C. roenbergensis*	19.1	20
OLV	Large DNA viruses	Phototropic algea	26.4	26
Sputnik 2	Lentille virus	*A. polyphaga*	18.3	20
YSLV1-4	Phycodna- or Mimivirus	Microalgea	27.8	26
ALM	Mimivirus?	Phagotrophic protozoan?	17.8	22

Figure 6 Overview over virophages of NCLDV. A. Unrooted phylogenetic tree of DNA packaging ATPase (45). **B.** Features of virophages (adapted from (45)).

EVOLUTION OF GIANT VIRUSES

After the discovery of giant viruses the question about their evolutionary origin arose. Different features of giant viruses lead to the question, whether giant viruses might form a fourth domain of life. Unlike the textbook definition of a virus, the giants are non-filterable agents. Giant viruses are also larger in particle and genome size than the cells of various bacteria or archaea, parasitic or free-living. Pandoravirus and Pithovirus show an uncommon, asymmetric virion structure. Also genetically, the giant viruses show some cell-like features. Giant viruses harbor genes with no detectable homologs and unknown sources, some of which are encoding components of the translation and core transcription machinery. This led to the idea that giant viruses might have to be incorporated in the tree of life as fourth domain. A maximum likelihood reconstruction of gene gain and loss events during evolution of giant viruses indicates that each family evolved from viruses with smaller genomes. Recent data suggests an accordion-like model of giant virus evolution which involves successive steps of gene gain and gene losses (46) Phylogenetic analysis of 13 genes involved in translation, which are also widely presented in at least 2 other domains of life showed, that 12 out 13 genes are separated from the eukaryotic root by multiple edges. Also, these genes are often polyphyletic amongst different families of giant viruses suggesting that the universal cellular genes of giant viruses were acquired by giant viruses from their eukaryotic hosts at different stages of evolution (16). Numerous genes of giant viruses are ORFans. These genes do not have any homologs, apart from genes in closely related strains, others have only very few homologs. One study analyzing 1’292 phylogenetic trees not including the core genes of NCLDV for origin interference found that the viral genes are primarily of eukaryotic origin. The variation between different families of giant viruses was found to be huge (Figure 7). This also indicates a distinct evolutionary history for giant viruses and contradicts the fourth domain hypothesis.

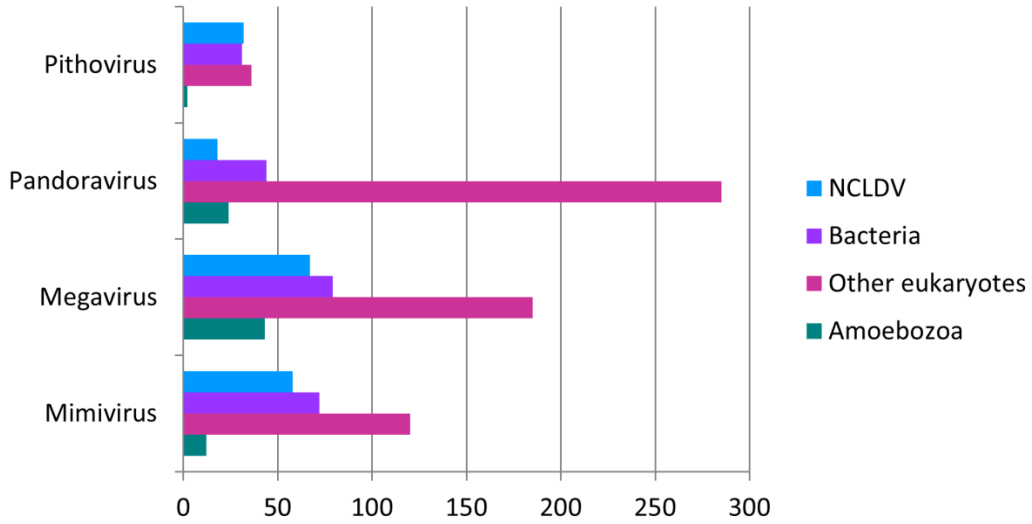


Figure 7 Phylogenetic breakdown of giant virus genes (adapted from (16)). Analysis of 1'292 phylogenetic trees of giant virus genes (excluding ORFans and genes with only few homologs and/or limited similarities to the detectable homologs) showing the evolutionary origin of giant virus genes.

POTENTIAL PATHOGENICITY OF GIANT VIRUSES IN HUMANS

After the discovery of Mimiviridae and other giant viruses all over the world from all kinds of sources it became obvious that these viruses might be a potential danger for humans. An analysis of the published literature in 2014 concludes that Mimiviridae and Marseilleviridae are potential agents of community-acquired and healthcare associated pneumonia. Infection with these viruses seems to be involved in poor outcome in patients in intensive care units. The exact mechanism underlying the induction of pneumonia by mimivirus remains unclear (47). Another study showed that mimivirus is taken up by professional macrophages, suggesting that alveolar macrophages could be a potential target for the propagation of mimivirus causing pneumonia (48). Several studies were performed with the aim to isolate giant viruses from patient's stool samples. Marseillevirus was found in stool from a healthy Senegalese, mimivirus in a

Tunisian patient with pneumonia, and Shan virus in another Tunisian patient with pneumonia. In case of Shan virus 1'605 stool samples were analyzed and one probe was positive indicating that giant viruses are rare in human stool (49). Our group showed that the exposure to mimivirus collagen L71 promotes rheumatoid arthritis in mice (50). Also, an analysis of human sera for anti mimivirus L71 antibodies found 22% positive samples in rheumatoid arthritis patients in comparison to 6% positive samples in healthy subjects (50). Given the wide distribution of Mimiviridae/ Marseilleviridae one can easily assume that humans come in contact with the viruses throughout their lives. Mimiviridae and Marseilleviridae, however, do not seem to be a human pathogen in a majority of cases.

PROTEIN GLYCOSYLATION

GENERAL CONCEPTS OF EUKARYOTIC PROTEIN GLYCOSYLATION

Most plasma-membrane and secretory proteins are post-translationally modified with carbohydrate chains. Glycosylation is an important chemical modification and plays various biological roles e.g. in mediating cell-cell interactions, protein folding and many more. Protein glycosylation mainly takes places in the ER and in the Golgi apparatus. There are two important types of carbohydrate modifications of proteins resulting in the formation of glycoproteins: N-glycosylation and O-glycosylation (Figure 8).

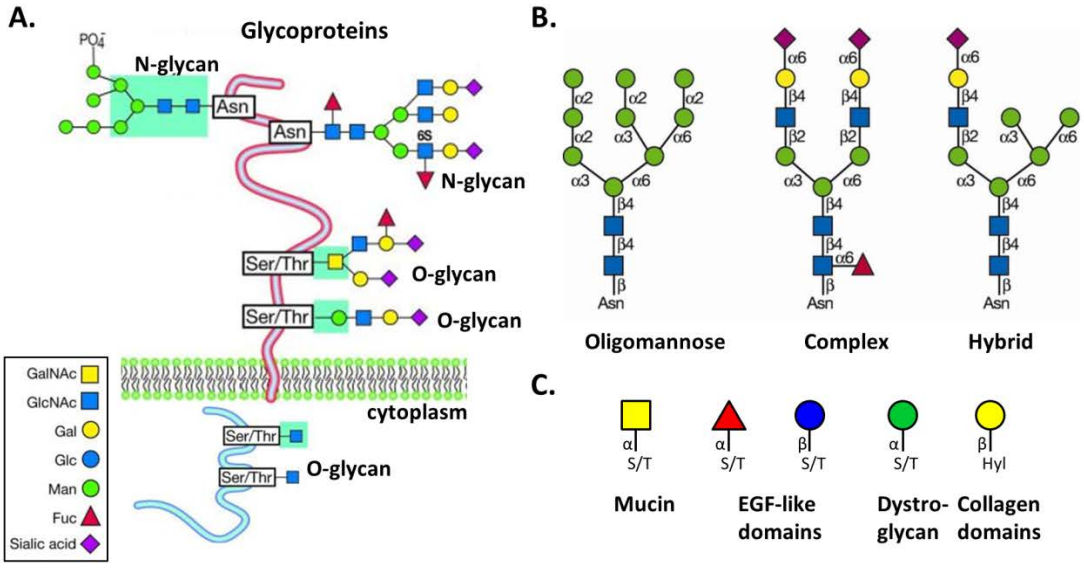


Figure 8 N- and O-glycosylation in animals. A. Examples for N- and O-glycosylation in animals (adapted from (51)). **B.** The three general types of N-glycosylation in a mature glycoprotein: oligomannose, complex, and hybrid, each with the common core $Man_3GlcNAc_2Asn$. **C.** Different types of O-glycosylation found in eukaryotes. Asn= asparagine S/ Ser= serine, T/ Thr= threonine, Hyl= hydroxylysine.

In eukaryotes N-glycosylation takes place in the ER. On a consensus sequence motif, namely Asn-X-Ser/Thr (where X is any amino acid but proline), a large glycan core is

added en bloc from the lipid carrier dolichol phosphate. After proper folding the protein can exit the ER towards the Golgi. In the Golgi the N-glycan undergoes various trimming and elongation steps to form a vast variety of different oligosaccharides (Figure 8B) (52). The synthesis of O-glycans is initiated by the addition of a single monosaccharide followed by stepwise extension of the glycan. The elongation of O-glycans mainly takes place in the Golgi apparatus. Many cytoplasmic and nuclear proteins that do not enter the secretory pathway gain their O-glycans in the cytoplasm. In O-glycosylation the sugar is added to the oxygen atom of serine, threonine, hydroxylysine, or hydroxyproline of a fully folded protein. There are different types of core structures in O-glycosylation forming different types of glycans with various functions (Figure 8C) (53). Glycosylation is performed by a specific class of enzymes, the glycosyltransferase. Glycosyltransferases are a prevalent enzyme type, found through-out all kingdoms of life. Glycosyltransferases represent about 1-2% of genomes. To date, more than 245'000 genes are classified as glycosyltransferases on the Carbohydrate Active enZYme database (CAZY). They are grouped within 97 distinct, sequence-based enzyme families (EC 2.4.1.x, 2.4.2.x, 2.4.99.x, 3.1.3.x, 3.2.1.x, 3.5.1.x, 5.4.99.x) (54, 55). If a glycosyltransferase shares similarity in at least 100 amino acids in two different stretches to members of one glycosyltransferase family but not to other families this glycosyltransferase is assigned to that specific family with the same predicted fold (56). More than 90% of the sequences annotated on CAZY are uncharacterized ORFs. Glycosyltransferases are defined as enzymes that catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules forming a glycosidic bond. Usually, glycosyltransferases are highly specific in regard to substrate, acceptor, and the glycosidic bond they form. Glycosyltransferases can be inverting or retaining depending on the stereochemistry of their substrates and products (Figure 9A). Given the large number of sequences it is striking that two types of active folds, termed GT-A and GT-B, (and variants thereof) have been identified for glycosyltransferases using

nucleotide sugars as substrates (Figure 9B) (57). The GT-A motif consists of a single domain with the first 120 amino acids forming a Rossmann fold. Rossmann folds are usually found in proteins binding nucleotides. Within the GT-A motif the Rossmann fold is responsible for binding the donor sugar nucleotide. With the exception of one enzyme, all glycosyltransferases with a GT-A domain possess a DXD motif which is responsible for the coordination of a metal ion. The GT-B fold consists of two domains separated by a cleft which is responsible for acceptor binding. Both domains show elements similar to Rossmann folds, but it is the carboxy terminal domain which is responsible for binding of the donor nucleotide sugar. GT-B fold containing enzymes do not have a DXD motif and are independent of metal ions (58).

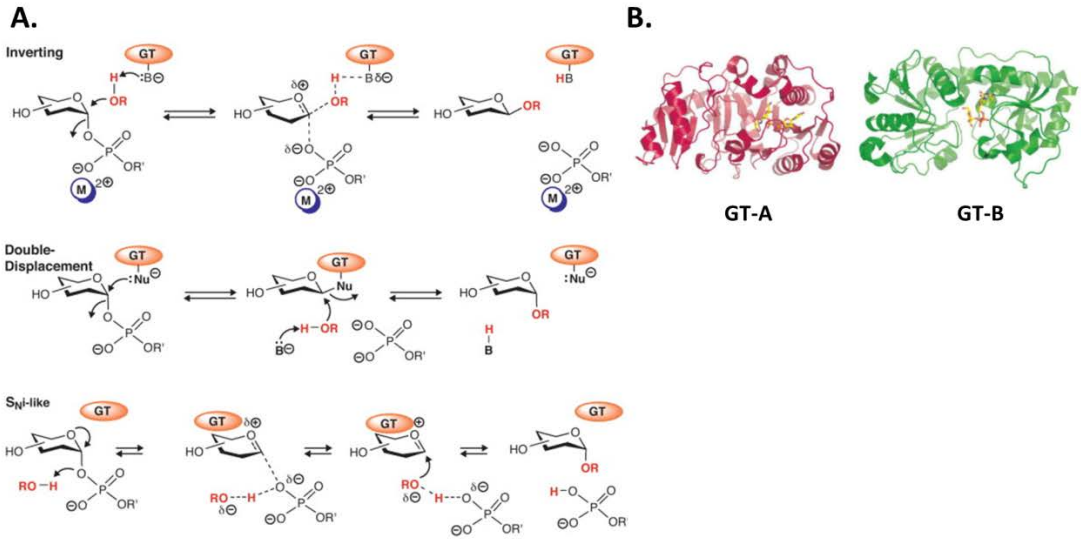


Figure 9 A. Proposed reaction mechanisms for inverting and retaining glycosyltransferases (GTs). Single displacement mechanism in an inverting glycosyltransferase reaction forming an oxocarbenium-ion in the transition state. The OH-group of the acceptor (HOR) is deprotonated by a catalytic amino acid. The negative charge on the phosphate group can be stabilized by a metal ion (M^{2+}) (GT-A enzymes) or by positive amino acids or a helix dipole (GT-B enzymes). The two proposed mechanisms for retaining enzymes: double displacement and S_Ni-like. Two

S_N2 reactions are involved in the double displacement mechanism. First the anomeric center of the donor sugar nucleotide is attacked resulting in a covalent glycosyl-enzyme intermediate. This intermediate is attacked by a hydroxyl-group of the acceptor. In the S_Ni mechanism an interaction between the departing phosphate and the acceptor nucleophile leads to the formation of an oxocarbenium ion that is stabilized by the glycosyltransferase. ROH = acceptor substrate. R' = donor sugar nucleotide (57). **B. Ribbon diagrams of representative GT-A and GT-B folds.** GT-A: rabbit β1-2-N-acetylglucosaminyltransferase I (PDB ID 1foa), GT-B: T4 phage β-glucosyltransferase (PDB ID 1j39). Bound nucleotide sugar shown in stick representation (58).

No correlation exists between the fold and reaction stereochemistry. Both inverting and retaining enzymes are known for GT-A and GT-B enzymes. Inverting glycosyl-transferase reactions are suggested to happen in a single displacement S_N2 reaction forming an intermediate oxocarbenium-ion. The negative charge of the leaving phosphate group can be stabilized by a metal ion in case of GT-A enzymes or by positive amino acids or a helix dipole in case of GT-B enzymes. For retaining enzymes two mechanisms have been proposed. The double displacement mechanism involves two S_N2 reactions. First the anomeric center of the donor sugar nucleotide is attacked resulting in a covalent glycosyl-enzyme intermediate. This intermediate is attacked by a hydroxyl-group of the acceptor. In the S_Ni mechanism an interaction between the departing phosphate and the acceptor nucleophile leads to the formation of an oxocarbenium-ion that is stabilized by the glycosyltransferase (Figure 9A) (57). In regard to kinetics, many glycosyltransferases have been shown to work in a Bi Bi sequential kinetic way. The donor sugar nucleotide binds before the acceptor substrate and the glycosylated acceptor is released before the nucleotide di- or monophosphate. This kind of kinetic goes in hand with the structural model in which the active site forms a deep pocket with the nucleotide donor sugar at the bottom and the acceptor substrate stacked on top (58). Usually, the donor sugar nucleotide binds the glycosyltransferase first inducing a conformational change. The glycosyltransferase is forming a lid over the

donor sugar nucleotide facilitating the binding of the acceptor substrate. This lid is formed by an internal disordered loop that becomes ordered when the donor sugar nucleotide is bound. A change in orientation and conformation of the now ordered loop facilitates the binding of the acceptor substrate and catalysis (56).

VIRUS GLYCOSYLATION

Although virus genomes usually do not encode any glycosyltransferases, structural proteins of viruses are often modified by both N- and O-glycosylation. Virus glycoproteins play diverse roles in the lifecycle of viruses. Virus glycoproteins can be important for host-virion binding, for entry; virus glycoproteins can serve as localization signals or protect virus particles from degradation and host immune response by forming a sugar shell masking the protein surface (Table 3). Glycosylation of virus envelope or surface proteins can promote proper folding and trafficking of virus proteins via the host's chaperones and folding factors. Virus glycoproteins are often complex and show dynamic glycosylation patterns: viral evolution due to high mutation rates allows the adding and deletion of glycosylation sites. A well-studied example of a virus protein is influenza hemagglutinin (HA). The surface glycoprotein HA influences various processes like receptor binding, infectivity, virus release, and neurovirulence. HA allows influenza virus to attach and enter its host cell through sialylated receptors. Upon binding HA is cleaved by a cellular protease. Both binding and cleavage of HA are influenced by the glycosylation of the protein. Also, the release of influenza virus from its host cell mediated by neuraminidase is influenced by the degree of HA glycosylation (59). Another famous example for virus glycoproteins is the envelope protein of HIV-1, gp120. Gp120 is one of the most heavily glycosylated proteins found in nature. A large number of high-mannose (Man) N-glycans decorate the protein. Gp120 binds to CD4 on the host cell surface. Gp120 is crucial for virus survival and immune evasion. A change in gp120 glycosylation alters the sensitivity of neutralizing antibodies

against HIV. The changing glycan shield of HIV-1 represents a mechanism for viral persistence despite an increasing antibody repertoire in the host (59).

Table 3 Examples for viral protein glycan structure and function (adapted from (59))

Virus Protein	Glycan Structure	Role
Influenza HA	High Man, GlcNAc	Attachment, release, glycan shield?
HIV-1 gp120	High Man, hybrid, sialic acid	Attachment, glycan shield
Hepatitis C virus E1, E2	High Man	Infectivity, entry
West Nile virus E	High Man, fucose	Replication, neuroinvasion
SARS-Corona virus S, M	High Man, hybrid, GalNAc	Assembly, attachment
Ebola GP	High Man	Infectivity

To achieve glycosylation the vast majority of viruses utilize host-encoded glycosyl-transferases located in the ER and Golgi apparatus (60). After folding and glycosylation, virus glycoproteins are usually transported to virus specific regions of host membrane by the sorting and transport mechanisms of the host cells (61). This means that the viral glycan pattern is host specific (Figure 10). The only strategies a virus can apply to alter its glycosylation pattern is either to infect a different cell type, affect the expression of host glycosyltransferases, or to change its own glycosylation sites through mutations.

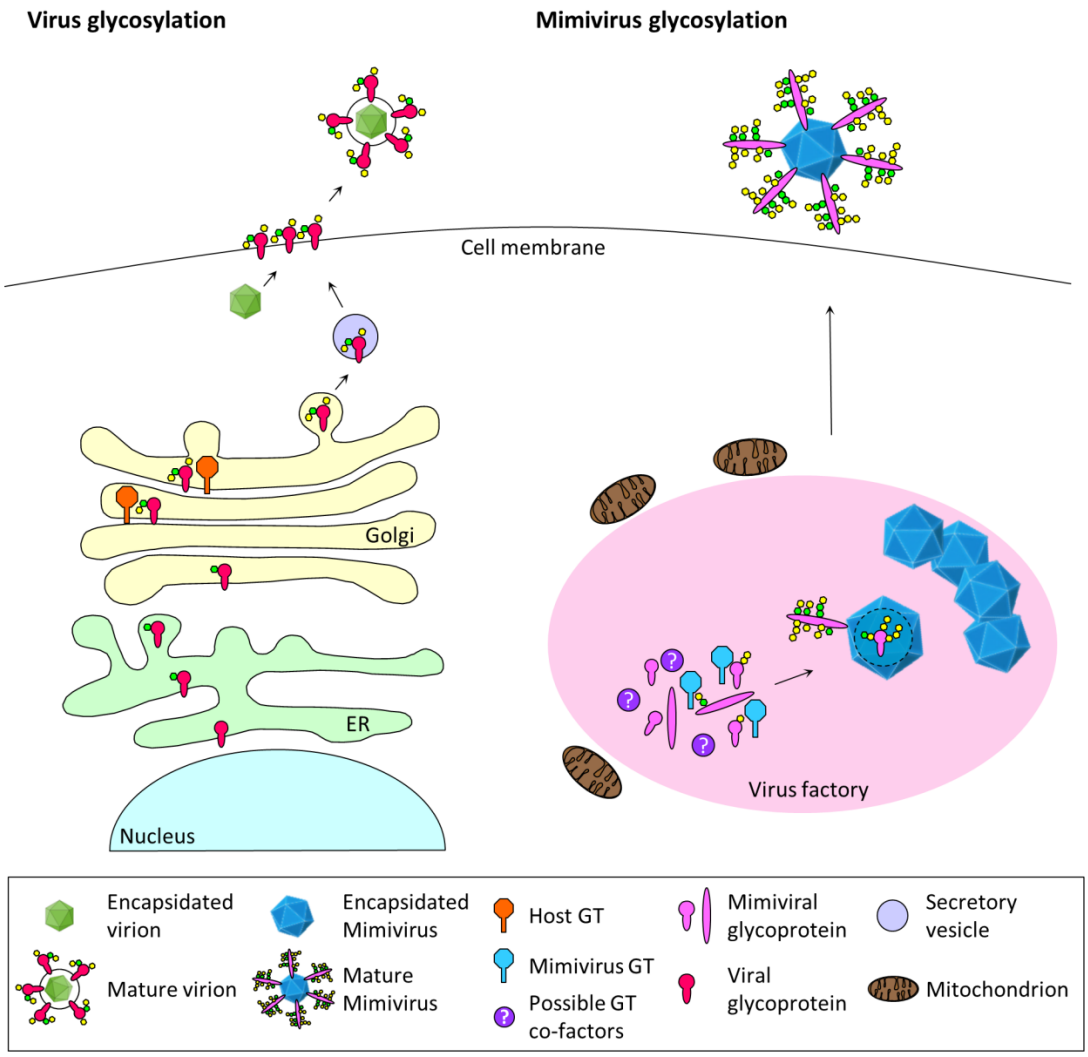


Figure 10 Schematic representation of virus glycosylation and mimivirus glycosylation. Glycoproteins of most viruses pass through the glycosylation and secretory machinery of their host. Therefore the glycoproteins carry a host specific glycan pattern. Mimivirus encodes its own glycosyltransferases and it is highly likely that glycosylation takes place in the virus factory independent of the host glycosylation machinery. Thus, mimivirus glycoproteins show a virus specific glycosylation pattern with unusual monosaccharides and linkages.

Recent findings show that few viruses encode their own glycosyltransferases (Table 4) (62). The diverse functions of these virus encoded glycosyltransferases illustrate several fascinating aspects of virology. For instance T2, T4, and T6 bacteriophages have the ability to avoid host anti-viral mechanisms by glycosylation of their DNA and thus protection from host-derived endonucleases (63). Other bacteriophages confer virulence factors in the bacteria they invade. This mechanism plays a crucial role in the development of new serotypes.

Table 4 Glycosyltransferases encoded by viruses (adapted from (63) and (29))

Virus	Glycosyltransferase	Biochemical Effect
Bacteriophages		
T2, T4, T6	α- and β-glucosyl-transferases	Glycosylation of viral DNA: protection from host endonuclease
Serotype converting phage	Specific for each bacteriophage	Conversion of bacterial O-antigen: inhibition of cell superinfection Host-cell serotype conversion
Baculoviruses (all except XcGV and PhopGV)	Ecdysteroid transferase	Glycosylation of ecdysteroid hormone: Inhibition of insect moulting and pupation
Chordopoxvirus	α-2,3-sialyl-transferase	Post-translational modification of SERP-1
Herpesviruses		
Rhadinovirus	Core 2 β-1,6-N-acetylglucosaminyl-transferase mucin type	Post-translational modification of structural proteins
Chloroviruses		
Paramecium bursaria chlorella virus	ORFs *a64r, a111r, a114r, a222-226r, a328l, a473l and a546l*	Post-translational modifications of the major capsid protein Vp54
	Hyaluronan synthase	Synthesis of a dense hyaluronan network on infected cell surface

Other bacteriophages confer virulence factors in the bacteria they invade. This mechanism plays a crucial role in the development of new serotypes. Other viruses have the ability to alter their host's metabolism. For example, most baculoviruses inhibit their host's moulting and pupation by glycosylating the insect hormone ecdysteroid. The described virus encoded glycosyltransferases are embedded in the host glycosylation machinery located in the ER and Golgi apparatus (63).

Chloroviruses, members of Phycodnaviridae, use a different system of glycosylation. The most prominent member of chloroviruses, PBCV-1, encodes at least five putative glycosyltransferases as well as enzymes involved in sugar metabolism and synthesis of polysaccharides. PBCV-1 particles assemble in so-called virus-assembly centers and glycosylation of virus proteins takes places in these centers independent of the host glycosylation machinery (29). The major capsid protein Vp54 of PBCV-1 carries a virus specific N-glycan (29). This N-glycan is not attached to the canonical sequon and the glycan structure is atypical and more reminiscent of bacterial and archaeal glycans (64). The mimivirus VF is a similar intracellular structure for replication and assembly and the putative mimivirus glycosyltransferases lack ER signaling sequences. Hence any glycosylation occurring in mimivirus is likely independent of AP (Figure 10).

GLYCOSYLATION IN MIMIVIRUS

Glycan content and O-glycosylation

The monosaccharide composition of mimivirus was analyzed in the past using two different approaches: gas chromatography-mass spectrometry (GC-MS) for analysis of the glycans on the surface fibers and high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) for analysis of the total monosaccharide content. The GC-MS study identifies rhamnose (Rha), glucose (Glc), and N-acetyl-glucosamine (GlcNAc) as main components of mimivirus glycans. Low amounts of ribose, arabinose, xylose (Xyl), Man, and galactose (Gal) are also found.

In addition to the known sugars two unusual sugars are found, one being viosamine (Vio) and the other suggested to being 3-O-methyl-Vio (65). The study using HPAEC-PAD identifies the following monosaccharides as components of mimivirus glycans: Rha, glucosamine (GlcN), Glc with a molar ratio of 1.9:2.6:1, and small amounts of fucose (Fuc) and Xyl. An unidentified monosaccharide is assigned to be methyl-Vio (Figure 11A) (66). The same study finds different groups of mimivirus O-glycans. Most mimivirus O-glycans are linear hexose polymer sequences. These linear O-glycans can be grouped according to their reducing end residues: di-methylated deoxy-hexose (di-methyl-Rha), methylated deoxyhexosamine (methyl-Vio), di-methylated hexose (di-methyl-Glc), hexose (Glc), and methylated hexosamine (methyl-GlcN). The oligohexoses correspond to Glc polymers. Mimivirus also has branched O-glycan structures. The branched O-glycans have a hexose (2-, 6-substituted glucitol) at the reducing end and terminal N-acetylhexosamine (residues with methyl modifications methyl-GlcNAc) and a terminal pentose residue (Xyl) in 3- or 4-linkage to N-acetylhexosamine (GlcNAc) (Figure 11B).

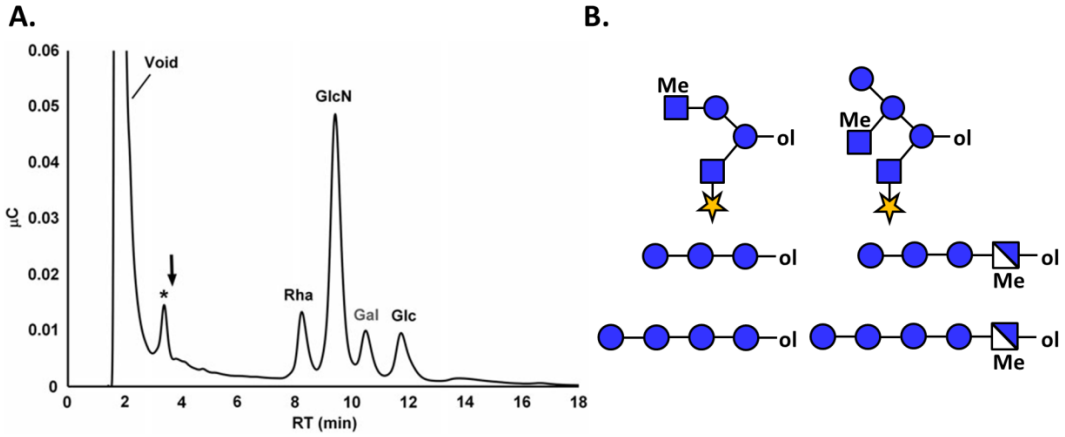


Figure 11 Monosaccharide analysis of mimivirus proteins (A) and schematic representation of selected mimivirus O-glycans (B). A. mimivirus protein extract was chromatographed by

HPAEC-PAD: Gal was used as internal standard for quantification. Black arrow: elution position of Vio. Asterisk: unidentified peak. **B.** Selected branched (top) and linear (middle and bottom) mimivirus O-glycans. Circle: Glc; square: GlcNAc; square with crossline: GlcN; star: Xyl; Me: methyl; ol: reducing end.

Nucleotide-diphospho-sugar synthesizing enzymes

The mimivirus genome encodes several genes involved in the synthesis of nucleotide-diphospho-sugars. In 2010 two enzymes were identified that are involved in the synthesis of UDP-Rha. R141 encodes a UDP-D-Glc-4, 6-dehydratase and L780 encodes a bifunctional UDP-4-keto-6-deoxy-D-Glc 3, 5-epimerase/4-reductase. The UDP-L-Rha synthesis pathways and enzyme properties are similar as those described for plants (67). Mimivirus L136 was identified as PLP-dependent sugar aminotransferase. It catalyzes the formation of UDP-Vio (4-amino-4, 6-dideoxy-D-Glc) from UDP-4-keto-6-deoxy-D-Glc (65). The synthesis of UDP-GlcNAc in mimivirus involves three proteins, namely L619, a glutamine-fructose-6-phosphate transaminase, L316, a glucosamine-6-phosphate N-acetyltransferase, and R689, a UDP-GlcNAc pyrophosphorylase. The synthesis of UDP-GlcNAc follows a eukaryotic-like strategy, but also shares some properties with the prokaryotic pathway (68). The genes described above are expressed at the intermediate to late stage of mimivirus infection, which would go in hand with their possible role in posttranslational modification of structural proteins and the fibers covering the surface of mimivirus. Indeed, the analysis of defibered viral particles finds a significant reduction of all monosaccharides, with a most pronounced effect on Rha, Vio, and GlcNAc, which were reduced to about 10% (65).

Glycosyltransferases

Within the mimiviral genome, eleven genes, namely L137, L138, L142, L193, L230, L373, R139, R363, R654, R655, and R707, are annotated as putative glycosyltransferases on CAZY (54, 55). The function of one of the genes, namely L230, was described in 2009. Mimivirus L230 is a bifunctional enzyme hydroxylating lysine and glycosylating the

resulting Hyl in native mimivirus collagen acceptor substrates. The glycosylating domain resides in the N-terminal domain and is important for the transfer of UDP-Glc to polypeptide Hyl (69). Apart from L230 nothing is known about the other glycosyl-transferases in mimivirus to date.

Knowing that mimivirus encodes its own nucleotide sugar synthesis machinery as well as at least one functional glycosyltransferase, which is responsible for glycosylation of mimivirus protein, we aim to analyze the functionality of other putative mimivirus glycosyltransferases.

REFERENCES

1. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, et al. A giant virus in amoebae. Science. 2003;299(5615):2033. Epub 2003/03/29.

2. Van Etten JL, Lane LC, Dunigan DD. DNA viruses: the really big ones (giruses). Annual review of microbiology. 2010;64:83-99. Epub 2010/08/10.

3. Kuznetsov YG, Xiao C, Sun S, Raoult D, Rossmann M, McPherson A. Atomic force microscopy investigation of the giant mimivirus. Virology. 2010;404(1):127-37. Epub 2010/06/17.

4. Eric Ghigo JK, Oham Lien, Lucas Pelkmans, Christian Capo, Jean-Louis Mege, Didier Raoult. Ameobal Pathogen Mimivirus Infects Macrophages through Phagocytosis. PloS Pathogens. 2008.

5. Klose T, Kuznetsov YG, Xiao C, Sun S, McPherson A, Rossmann MG. The three-dimensional structure of Mimivirus. Intervirology. 2010;53(5):268-73. Epub 2010/06/17.

6. Legendre M, Santini S, Rico A, Abergel C, Claverie JM. Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. Virology journal. 2011;8:99. Epub 2011/03/08.

7. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. Science. 2004;306(5700):1344-50. Epub 2004/10/16.

8. Wessner DR. Discovery of the Giant Mimivirus. Nature Education. 2010;3(9):61.

9. Korn ED, Weisman RA. Phagocytosis of latex beads by Acanthamoeba. II. Electron microscopic study of the initial events. The Journal of cell biology. 1967;34(1):219-27. Epub 1967/07/01.

10. Zauberman N, Mutsafi Y, Halevy DB, Shimoni E, Klein E, Xiao C, et al. Distinct DNA exit and packaging portals in the virus Acanthamoeba polyphaga mimivirus. PLoS biology. 2008;6(5):e114. Epub 2008/05/16.

11. Mutsafi Y, Fridmann-Sirkis Y, Milrot E, Hevroni L, Minsky A. Infection cycles of large DNA viruses: Emerging themes and underlying questions. Virology. 2014. Epub 2014/07/06.

12. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(42):17486-91. Epub 2011/10/12.

13. Colson P, Yutin N, Shabalina SA, Robert C, Fournous G, La Scola B, et al. Viruses with more than 1,000 genes: Mamavirus, a new Acanthamoeba polyphaga mimivirus strain, and reannotation of Mimivirus genes. Genome biology and evolution. 2011;3:737-42. Epub 2011/06/28.

14. Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, et al. Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome biology and evolution. 2012;4(12):1324-30. Epub 2012/12/12.

15. Fischer MG, Allen MJ, Wilson WH, Suttle CA. Giant virus with a remarkable complement of genes infects marine zooplankton. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(45):19508-13. Epub 2010/10/27.

16. Yutin N, Wolf YI, Koonin EV. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology. 2014. Epub 2014/07/22.

17. Colson P, De Lamballerie X, Yutin N, Asgari S, Bigot Y, Bideshi DK, et al. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Archives of virology. 2013;158(12):2517-21. Epub 2013/07/03.

18. Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, et al. ViralZone: a knowledge resource to understand virus diversity. Nucleic acids research. 2011;39(Database issue):D576-82. Epub 2010/10/16.

19. Yutin N, Wolf YI, Raoult D, Koonin EV. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virology journal. 2009;6:223. Epub 2009/12/19.

20. Klose T, Rossmann MG. Structure of large dsDNA viruses. Biological chemistry. 2014;395(7-8):711-9. Epub 2014/07/09.

21. Koonin EV, Yutin N. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology. 2010;53(5):284-92. Epub 2010/06/17.

22. Philippe N, Legendre M, Doutre G, Coute Y, Poirot O, Lescot M, et al. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science. 2013;341(6143):281-6. Epub 2013/07/23.

23. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, et al. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(11):4274-9. Epub 2014/03/05.

24. Takamatsu HH, Denyer MS, Lacasta A, Stirling CM, Argilaguet JM, Netherton CL, et al. Cellular immunity in ASFV responses. Virus research. 2013;173(1):110-21. Epub 2012/12/04.

25. Federici BA, Bideshi DK, Tan Y, Spears T, Bigot Y. Ascoviruses: superb manipulators of apoptosis for viral replication and transmission. Current topics in microbiology and immunology. 2009;328:171-96. Epub 2009/02/17.

26. Chen G, Robert J. Antiviral immunity in amphibians. Viruses. 2011;3(11):2065-86. Epub 2011/12/14.

27. Wilson WH, Van Etten JL, Allen MJ. The Phycodnaviridae: the story of how tiny giants rule the world. Current topics in microbiology and immunology. 2009;328:1-42. Epub 2009/02/17.

28. Yamada T. Giant viruses in the environment: their origins and evolution. Current opinion in virology. 2011;1(1):58-62. Epub 2012/03/24.

29. Van Etten JL, Gurnon JR, Yanai-Balser GM, Dunigan DD, Graves MV. Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the

endoplasmic reticulum and Golgi. Biochimica et biophysica acta. 2010;1800(2):152-9. Epub 2009/08/06.

30. Schmidt FI, Bleck CK, Mercer J. Poxvirus host cell entry. Current opinion in virology. 2012;2(1):20-7. Epub 2012/03/24.

31. Smith KA. Smallpox: can we still learn from the journey to eradication? The Indian journal of medical research. 2013;137(5):895-9. Epub 2013/06/14.

32. Thomas V, Bertelli C, Collyn F, Casson N, Telenti A, Goesmann A, et al. Lausannevirus, a giant amoebal virus encoding histone doublets. Environmental microbiology. 2011;13(6):1454-66. Epub 2011/03/12.

33. Boyer M, Yutin N, Pagnier I, Barrassi L, Fournous G, Espinosa L, et al. Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(51):21848-53. Epub 2009/12/17.

34. Aherfi S, La Scola B, Pagnier I, Raoult D, Colson P. The expanding family Marseilleviridae. Virology. 2014. Epub 2014/08/12.

35. Abrahao JS, Dornas FP, Silva LC, Almeida GM, Boratto PV, Colson P, et al. Acanthamoeba polyphaga mimivirus and other giant viruses: an open field to outstanding discoveries. Virology journal. 2014;11:120. Epub 2014/07/01.

36. Campos RK, Boratto PV, Assis FL, Aguiar ER, Silva LC, Albarnaz JD, et al. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. Virology journal. 2014;11:95. Epub 2014/06/03.

37. Yoosuf N, Pagnier I, Fournous G, Robert C, Raoult D, La Scola B, et al. Draft genome sequences of Terra1 and Terra2 viruses, new members of the family Mimiviridae isolated from soil. Virology. 2014;452-453:125-32. Epub 2014/03/13.

38. Pagnier I, Reteno DG, Saadi H, Boughalmi M, Gaia M, Slimani M, et al. A decade of improvements in Mimiviridae and Marseilleviridae isolation from amoeba. Intervirology. 2013;56(6):354-63. Epub 2013/10/26.

39. La Scola B, Campocasso A, N'Dong R, Fournous G, Barrassi L, Flaudrops C, et al. Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. Intervirology. 2010;53(5):344-53. Epub 2010/06/17.

40. La Scola B, Barrassi L, Raoult D. Isolation of new fastidious alpha Proteobacteria and Afipia felis from hospital water supplies by direct plating and amoebal co-culture procedures. FEMS microbiology ecology. 2000;34(2):129-37. Epub 2000/12/05.

41. Gaia M, Pagnier I, Campocasso A, Fournous G, Raoult D, La Scola B. Broad spectrum of mimiviridae virophage allows its isolation using a mimivirus reporter. PloS one. 2013;8(4):e61912. Epub 2013/04/19.

42. Sun S, La Scola B, Bowman VD, Ryan CM, Whitelegge JP, Raoult D, et al. Structural studies of the Sputnik virophage. Journal of virology. 2010;84(2):894-7. Epub 2009/11/06.

43. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, et al. The virophage as a unique parasite of the giant mimivirus. Nature. 2008;455(7209):100-4. Epub 2008/08/12.

44. Fischer MG, Suttle CA. A virophage at the origin of large DNA transposons. Science. 2011;332(6026):231-4. Epub 2011/03/10.

45. Zhou J, Zhang W, Yan S, Xiao J, Zhang Y, Li B, et al. Diversity of virophages in metagenomic data sets. Journal of virology. 2013;87(8):4225-36. Epub 2013/02/15.

46. Filee J. Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution. Frontiers in microbiology. 2015;6:593. Epub 2015/07/03.

47. Kutikhin AG, Yuzhalin AE, Brusina EB. Mimiviridae, Marseilleviridae, and virophages as emerging human pathogens causing healthcare-associated infections. GMS hygiene and infection control. 2014;9(2):Doc16. Epub 2014/08/26.

48. Ghigo E, Kartenbeck J, Lien P, Pelkmans L, Capo C, Mege JL, et al. Ameobal pathogen mimivirus infects macrophages through phagocytosis. PLoS Pathog. 2008;4(6):e1000087. Epub 2008/06/14.

49. Saadi H, Reteno DG, Colson P, Aherfi S, Minodier P, Pagnier I, et al. Shan virus: a new mimivirus isolated from the stool of a Tunisian patient with pneumonia. Intervirology. 2013;56(6):424-9. Epub 2013/10/26.

50. Shah N, Hulsmeier AJ, Hochhold N, Neidhart M, Gay S, Hennet T. Exposure to mimivirus collagen promotes arthritis. Journal of virology. 2014;88(2):838-45. Epub 2013/11/01.

51. Varki A, Sharon N. Historical Background and Overview. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

52. Stanley P, Schachter H, Taniguchi N. N-Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

53. Brockhausen I, Schachter H, Stanley P. O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

54. Campbell JA, Davies GJ, Bulone V, Henrissat B. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. The Biochemical journal. 1997;326 (Pt 3):929-39. Epub 1997/10/23.

55. Coutinho PM, Deleury E, Davies GJ, Henrissat B. An evolving hierarchical family classification for glycosyltransferases. Journal of molecular biology. 2003;328(2):307-17. Epub 2003/04/15.

56. Brockhausen I. Crossroads between Bacterial and Mammalian Glycosyltransferases. Frontiers in immunology. 2014;5:492. Epub 2014/11/05.

57. Breton C, Fournel-Gigleux S, Palcic MM. Recent structures, evolution and mechanisms of glycosyltransferases. Current opinion in structural biology. 2012;22(5):540-9. Epub 2012/07/24.

58. Rini J, Esko J, Varki A. Glycosyltransferases and Glycan-processing Enzymes. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

59. Vigerust DJ, Shepherd VL. Virus glycosylation: role in virulence and immune interactions. Trends in microbiology. 2007;15(5):211-8. Epub 2007/04/03.

60. Flint SJE, L.W.; Racaniello, V.R.; Skalka, A.M. Principles of Virology, third edition. Washington, DC: ASM Press; 2009.

61. Doms RW, Lamb RA, Rose JK, Helenius A. Folding and assembly of viral membrane proteins. Virology. 1993;193(2):545-62. Epub 1993/04/01.

62. Rohozinski J, Girton LE, Van Etten JL. Chlorella viruses contain linear nonpermuted double-stranded DNA genomes with covalently closed hairpin ends. Virology. 1989;168(2):363-9. Epub 1989/02/01.

63. Markine-Goriaynoff N, Gillet L, Van Etten JL, Korres H, Verma N, Vanderplasschen A. Glycosyltransferases encoded by viruses. The Journal of general virology. 2004;85(Pt 10):2741-54. Epub 2004/09/28.

64. De Castro C, Molinaro A, Piacente F, Gurnon JR, Sturiale L, Palmigiano A, et al. Structure of N-linked oligosaccharides attached to chlorovirus PBCV-1 major capsid protein reveals unusual class of complex N-glycans. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(34):13956-60. Epub 2013/08/07.

65. Piacente F, Marin M, Molinaro A, De Castro C, Seltzer V, Salis A, et al. The giant DNA virus Mimivirus encodes a pathway for the biosynthesis of the unusual sugar 4-amino-4,6-dideoxy-D-glucose (viosamine). The Journal of biological chemistry. 2011. Epub 2011/12/14.

66. Hulsmeier AJ, Hennet T. O-Linked glycosylation in Acanthamoeba polyphaga mimivirus. Glycobiology. 2014;24(8):703-14. Epub 2014/05/06.

67. Parakkottil Chothi M, Duncan GA, Armirotti A, Abergel C, Gurnon JR, Van Etten JL, et al. Identification of an L-rhamnose synthetic pathway in two nucleocytoplasmic large DNA viruses. Journal of virology. 2010;84(17):8829-38. Epub 2010/06/12.

68. Piacente F, Bernardi C, Marin M, Blanc G, Abergel C, Tonetti MG. Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNA virus Mimivirus. Glycobiology. 2014;24(1):51-61. Epub 2013/10/11.

69. Luther KB, Hulsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennet T. Mimivirus collagen is modified by a bifunctional lysyl hydroxylase and glycosyltransferase enzyme. The Journal of biological chemistry. 2011. Epub 2011/11/03.

MANUSCRIPT I: GLYCOGENIN-1 PARALOG IN GIANT MIMIVIRUS

Giant mimivirus R707 encodes a glycogenin paralog polymerizing glucose through alpha and beta glycosidic linkages

Biochemical Journal, in press

Anna J. Rommel[1], Andreas J. Hülsmeier[1], Simon Jurt[2], and Thierry Hennet[1]*

[1]Institute of Physiology, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland; [2]Institute of Organic Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

*Corresponding author
E-mail: thierry.hennet@uzh.ch
Phone: +41 44 635 50 80

Experiments were designed by AR and TH. All experiments except for MS and NMR experiments were done by AR. The manuscript was written by AR and TH.

ABSTRACT

Acanthamoeba polyphaga mimivirus is a giant virus encoding 1'262 genes among which many were previously thought to be exclusive to cellular life. For example, mimivirus genes encode enzymes involved in the biosynthesis of nucleotide sugars and putative glycosyltransferases. We identified in mimivirus a glycogenin-1 homologous gene encoded by the open reading frame R707. The R707 protein was found to be active as a polymerizing glucosyltransferase enzyme. Like glycogenin-1 R707 activity was divalent metal ion dependent and relied on an intact DXD motif. In contrast to glycogenin-1, R707 was however not self-glucosylating. Interestingly, the product of R707 catalysis featured α1,6, β1,6, and α1,4 glycosidic linkages. Mimivirus R707 is the first reported glycosyltransferase able to catalyze the formation of both α and β linkages. Mimivirus encoded glycans play a role in the infection of host amoeba. Co-infection of *Acanthamoeba* with mimivirus and amylose and chitin hydrolysate reduced the number of infected amoeba, thus supporting the importance of polysaccharide chains in the uptake of mimivirus by amoeba. The identification of a glycosyltransferase capable of forming α and β linkages underlines the peculiarity of mimivirus and enforces the concept of a host-independent glycosylation machinery in mimivirus.

SUMMARY STATEMENT

We have identified the mimivirus R707 protein as paralog to glycogenin-1 with polymerizing glucosyltransferase activity, catalyzing the formation of α1,6, β1,6, and α1,4 glycosidic linkages. Mimivirus R707 is the first reported glycosyltransferase capable of forming both α and β linkages.

KEYWORDS: giant virus / polysaccharide / glycogen / glycosylation / nuclear magnetic resonance

SHORT TITLE: Glycogenin-1 paralog in giant mimivirus

ABBREVIATIONS

CID-MS	collision-induced dissociation mass spectrometry
DQF-COSY	double quantum filtered correlated spectroscopy
ESI-MS	electrospray ionization mass spectrometry
Gal	galactose
GC-MS	gas-chromatography mass spectrometry
Glc	glucose
GlcNAc	N-acetylglucosamine
GYG1	glycogenin-1
HCD-MS	higher-energy collisional dissociation mass spectrometry
HexNAc	N-acetylhexosamine
HMBC	heteronuclear multiple bind correlation
HSQC	heteronuclear single quantum coherence
MALDI-MS	matrix-assisted laser desorption/ionization mass spectrometry
NMR	nuclear magnetic resonance
PBCV-1	*Paramecium bursaria* chlorella virus
pNP	4-nitrophenyl
Rha	rhamnose
ROESY	rotating frame Overhauser enhancement spectroscopy
TOCSY	two-dimensional nuclear magnetic resonance spectroscopy
UDP	uridine diphosphate
Vio	viosamine
Xyl	xylose

INTRODUCTION

The giant *Acanthamoeba polyphaga* mimivirus is a member of the expanding family of *Mimiviridae*, belonging to the monophyletic group of nucleocytoplasmic large DNA viruses. The genome of mimivirus consisting of 1.2 Mbp and encoding 1'262 genes has been described in 2003 (1). The icosahedral mimivirus particle is 400 nm in diameter and covered with 140 nm long glycosylated protein fibers (2). Many mimivirus genes are unusual for viruses as they code for genes typical of cellular life, such as tRNA synthases, DNA repair factors, chaperones, and metabolic enzymes (3). Interestingly mimivirus encodes multiple genes possibly composing a functional glycosylation machinery.

Virus glycoproteins are associated with host-virion binding, virus entry, as for example hepatitis C virus E1/E2, and host-immune evasion as HIV gp120 (4). Glycosylation of virus envelope or surface proteins can promote proper folding, trafficking, and stability of virus proteins, such as for example in influenza virus hemagglutinin (4). In contrast to giant viruses conventional viruses hijack the glycosylation machinery of their host and, with few examples, do not encode glycosyltransferases or other genes involved in glycosylation (4). A giant virus encoding its own glycosylation machinery is found in the family of Chloroviruses. *Paramecium bursaria* chlorella virus (PBCV-1), a member of *Phycodnaviridae* that infects unicellular green algae, encodes at least five putative glycosyltransferases as well as multiple enzymes involved in sugar metabolism (5). The major capsid protein Vp54 of PBCV-1 carries a virus specific N-glycan that is added to the protein independently of the host's ER and Golgi pathway (5). This N-glycan is not attached to the canonical sequon and the glycan structure is atypical and more reminiscent of bacterial and archaeal glycans (6).

In mimivirus the biosynthetic pathways of three nucleotide sugars are well characterized. Mimivirus synthesizes UDP-viosamine (Vio), UDP-rhamnose (Rha), and

UDP-N-acetylglucosamine (GlcNAc) independently of its host *Acanthamoeba polyphaga* (7-9). These nucleotide sugar donors reflect the monosaccharide composition of mimivirus, given that GlcNAc, Rha, and Vio are major carbohydrates isolated from mimivirus particles in addition to glucose (Glc) (8, 10). The analysis of the O-glycan content of mimivirus shows various glycan structures demanding for glycosyltransferases with different functionalities. Mimivirus proteins are glycosylated with linear O-glycans consisting of hexose polymers of variable length. Also branched O-glycans have been identified, which consist of a hexose at the reducing end, a terminal N-acetylhexosamine (HexNAc) residue on one branch and a terminal pentose residue linked to HexNAc on the second branch (10).

Analysis of the mimivirus genome suggests that in addition to the nucleotide-diphospho sugar synthesizing machinery mimivirus encodes at least 11 putative glycosyltransferases (3). These putative glycosyltransferases share sequence similarities with archaeal, bacterial, and eukaryotic glycosyltransferases. For example, the mimivirus L230 gene encodes a bi-functional collagen-modifying enzyme structurally similar to PLOD3 lysyl hydroxylase and GLT25D1 galactosyltransferase of animal origin (11). Another mimivirus protein, R707, is structurally related to glycogenin-1 (GYG1), which initiates glycogen biosynthesis in animals. Despite this similarity, R707 unlikely mediates glycogen formation in mimivirus, but may be involved in the assembly of polysaccharide decorating the virus capsid. The present study characterizes the glycosyltransferase activity of mimivirus R707 and the importance of polysaccharides in the infection of amoeba by mimivirus.

EXPERIMENTAL/ MATERIALS AND METHODS

Expression and purification of recombinant R707 protein – Recombinant R707 was expressed in *Escherichia coli* BL21 De3 (Novagen, Nottingham, UK) as His_{10}-fusion protein in pET16b (Merck, Billerica, MA). Viral DNA was purified using phenol/chloroform (Roth,

Karlsruhe, Germany). The DNA sequence corresponding to mimivirus R707 (RefSeq ID NC_014649.1) was amplified using the primers 5′-ATACTCGAGTCTTCCT-ATGCATATGTTACAG-3′ and 5′-ATAGGATCCTTAATAAGGTAGTTTAATGTCA-3′ containing XhoI and BamHI restriction sites, respectively (underlined bases). Site-directed mutagenesis of R707 was performed by QuickChange site-directed muta-genesis (12) using primers listed in table 1. Asn position 103 was substituted for Ser, Tyr 215 was substituted for Phe, and Lys 232 was substituted for Arg. For protein expression 100 ml lysogeny broth containing 100 μg/ml ampicillin (Sigma-Aldrich, Buchs, Switzerland) were inoculated with 1 ml of overnight culture. Protein expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (Biosolve, Dieuze, France) when bacteria reached OD_{600} value of 0.4. *E. coli* were grown at 16°C for 16 h, shaking, then pelleted at 6′000 x g, resuspended in 15 ml 500 mM NaCl, 20 mM Tris pH 7.4, 10% v/v glycine, 10 mM imidazole (MCAC10) and lysed using a French Press (Avestin, Mannheim, Germany). His_{10}-tagged R707 was purified from the soluble fraction using 100 μl Ni-Sepharose 6 Fast Flow beads (GE Healthcare, Glattbrugg, Switzerland). After binding to beads in MCAC10 at 4°C for 4 h, rotating, R707 was eluted in 400 mM imidazole (MCAC400). Purified R707 was stored at 4°C.

Enzymatic activity – Glycosyltransferase activity was measured in 20 mM Tris-HCl pH 7.4 containing 0.5 μl 20 μCi/ml UDP-[^{14}C]-sugar (Perkin-Elmer, Waltham, MA), 20 μM UDP-sugar, 10 mM $MnCl_2$ (Sigma-Aldrich), 1 mM DTT, 50 mM 4-nitrophenyl (pNP)-sugar in DMSO (Sigma-Aldrich) as acceptor substrate, and 10 μg of purified R707 protein. The UDP-sugars tested as donor substrates were UDP-Glc, UDP-GlcNAc, UDP-galactose (Gal), and UDP-xylose (Xyl). The pNP-sugars tested as acceptor substrates were pNP-Glc, pNP-Xyl, pNP-Gal, and pNP-Rha (Sigma-Aldrich). Assays were incubated at 35°C for 4 h, then stopped by addition of ice cold 500 μl H_2O, and purified over SepPak-C18 cartridges (Waters, Milford, MA). Samples were added to the cartridges, washed with 10 ml of H_2O, and eluted with 5 ml of methanol (Sigma-

Aldrich). Radioactivity was measured after addition of 10 ml of IRGASAFE scintillation liquid (Perkin-Elmer) in a Tri-Carb 2900TR scintillation counter (Perkin-Elmer). Self-glucosylation of R707 was measured as described above, but without addition of pNP-sugar acceptor substrate. After incubation assays were stopped by addition of 500 µl 5% trichloroacetic acid / 5% phosphotungstic acid and incubated on ice for 30 min. Precipitates were recovered on 1 µm glass filters (VWR, Radnor, PA) using a vacuum manifold and radioactivity was measured as described above.

Product analysis of R707 glycosyltransferase assay – About 200 pmol of glycosyltransferase assay product were subjected to MALDI-MS analysis. MALDI-MS were acquired essentially as described earlier (13). For NMR analysis assay products from 250 glycosyltransferase assays were pooled, dried, and resuspended in 300 µl DMSO. To remove spare pNP-α-Glc, products were purified by gel-filtration using a Superdex Peptide HR 10/30 (GE Healthcare) and 10% acetonitrile, 0.1% trifluoroacetic acid at a flow rate of 0.1 ml/min. Different product species were separated using a Hypersil C18 column (Thermo Scientific, Waltham, MA) under the following conditions: buffer A: 0.1% trifluoroacetic acid (Sigma-Aldrich), buffer B: 0.1% trifluoroacetic acid, 10% acetonitrile (Sigma-Aldrich), 30 min 97.5% buffer A/ 2.5% buffer B, 120 min gradient to 10% buffer B, flow rate: 0.5 ml/min. Product elution was monitored at 375 nm, and 2 min-fractions were collected. Peaks containing a sufficient amount of product were pooled, dried, and subjected to NMR analysis. Products were dissolved in 220 µl D_2O and transferred into 5 mm D_2O-matched Shigemi NMR tubes (Shigemi, Allison Park, PA). Structure elucidation of sample A and B is based on a set of 1D proton NMR, 2D ^{1}H-^{1}H double quantum filtered correlated spectroscopy (DQF-COSY), ^{1}H-^{1}H two-dimensional nuclear magnetic resonance spectroscopy (TOCSY) (100 ms mixing time), ^{1}H-^{1}H rotating frame Overhauser enhancement spectroscopy (ROESY) (300 ms mixing-time), ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) and ^{1}H-^{13}C heteronuclear multiple bind correlation (HMBC) experiments. Identification of sample C is based on

1D ^{1}H and 2D ^{1}H-^{13}C HSQC spectra that were also recorded for the reference compounds, namely pNP-α-maltose (Sigma-Aldrich). All spectra were collected at 298 K on a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenically cooled 5 mm TCI probe and externally referenced against 4,4-dimethyl-4-silapentane-1-sulfonate with a $^{13}C/^{1}H$ ratio of 0.25144953 used for the indirect ^{13}C referencing (14). Spectra were processed using the Bruker TopSpin 3.5 software suite. For linkage analysis of the products A, B, and C GC-MS experiments were done essentially as described previously (10). The initial oven temperature of 80°C was held for 2 min, increased by 8°C/min until 320°C and held for 8 min. In addition, the products A, B, and C were analyzed by negative-ion ESI-MS and ESI-HCD-MS/MS experiments were carried out on an Orbitrap Fusion mass spectrometer (Thermo Scientific), analogous to the negative ion ESI-CID-MS/MS method described recently (15). The glycosyltransferase reaction products were dissolved in 50 % acetonitrile, 2 mM NH_4HCO_3 (Sigma-Aldrich) at a concentration of 30 μM and 2 μl were injected to the mass spectrometer. Solvent 50 % acetonitrile, 2 mM NH_4HCO_3 was delivered by a Harvard syringe pump at a flow rate of 4 ml/min. The ion- transfer tube temperature was maintained at 320°C and the capillary voltage was maintained at – 2700 V. For MS/MS acquisition quadrupole isolation with a 2 m/z isolation width and HCD activation with 0 % collisional energy was used. Ions were detected in the Orbitrap at a resolution of 15'000 and a scan range of 50-500 m/z. The AGC target was set to 5×10^4 and maximal injection time to 100 ms.

Phagocytosis competition assay – Amoeba (5×10^3) were seeded in 96-well plates in peptone yeast glucose broth and infected at a multiplicity of 250 with purified mimivirus particles together with 1 to 10 μg of amylose, chitin hydrolysate, dextran, or mannan (Sigma-Aldrich). Polyvinyl alcohol (Sigma-Aldrich) was used as negative control. Amoeba were incubated with mimivirus and polymers at 28°C for 4.5 h, and then fixed in 4% paraformaldehyde. Intracellular DNA including virus factories was stained with DAPI for 30 min. Virus factories per total number of amoeba per view field

were quantified using an Axiovert 200M microscope (Zeiss, Oberkochen, Germany) and Fiji cell counter (16).

Statistical analysis - ANOVA with multiple comparisons was used and a P-value of $P < 0.05$ was considered as statistically significant. GraphPad Prism 6 (GraphPad Software, Inc., La Jolla, CA) was used for statistical analysis. Results are presented as mean ± SEM.

RESULTS

The mimivirus R707 protein was identified as putative glycosyltransferase based on sequence similarity to animal GYG1, which primes glycogen synthesis by initiated self-glucosylation (17). GYG1 transfers Glc from UDP-Glc to a conserved Tyr residue at position 195, then elongates this Glc residue with up to seven additional Glc units through α1,4-linkage, making GYG1 a retaining enzyme (17). R707 and GYG1 shared 26% sequence identity and 42% similarity. The mimivirus R707 protein included a DXD motif, which aligns with the same motif in animal GYG1 proteins (Figure 1A). The DXD motif is typically found in glycosyltransferases utilizing nucleotide-activated sugars as donor substrates, and is involved in the coordination of a divalent metal ion in the catalytic pocket of the enzyme (18). In GYG1 Leu position 8, Thr 10, Asn 11, Tyr 14, Asp 101, Ala 102, Asp 103, Asn 132, Gln 163, His 211, Gly 214, and Lys 217 are involved in the coordination of UDP-Glc. The homologous residues Asn 13, Tyr 16, Asp 101, Asp 103, Asn 150, His 226, and Lys 232 were also found to be conserved in mimivirus R707 (Figure 1B) (19). The pattern of sequence and motif similarity between GYG1 and R707 suggested a common Glc polymerizing activity for both proteins. Based on BLAST analysis the putative glycosyltransferase domain spanned nearly the entire R707 protein between amino acid position 4 and 265 (Figure 1C). A 3D-model comparison of R707 and GYG1 using PHYRE^2 and SWISS-MODEL shows strong conservation around the DXD motif and main secondary structures. The majority of the α-helices and β-sheets of GYG1 from residue 1-230 align with R707 (20, 21). Both proteins show structural

diversity towards the C-terminus as predicted by BLAST analysis. The R707 gene is expressed in the intermediate and late phase of mimivirus replication, indicating a possible involvement of R707 in virion maturation and fiber formation (22).

The R707 protein was expressed in *E. coli* as His_{10}-fusion protein, purified by affinity chromatography, and glycosyltransferase activity was tested. For product purification pNP-α-Glc and pNP-β-Glc were used as acceptor substrates. In presence of pNP-α-Glc R707 showed a specific activity of 19.6 pmol/min/mg. In presence of pNP-β-Glc R707 activity reached only 7.1 pmol/min/mg (Figure 2A). To test for a possible self-glucosylation activity of R707, we omitted the acceptor substrate from the enzymatic reaction. We could not detect any self-glucosylation activity on R707 (Figure 2A).

To assess the functionality of different residues of R707, we mutated amino acids in R707, which correspond to the essential residues Asp 103, Tyr 195, and Lys 217 in GYG1. The residue Asp 103 of R707 is part of the conserved DXD motif, Tyr 215 corresponds to the glucosylated Tyr residue of GYG1, and Lys 232 corresponds to a residue involved in the coordination of UDP-Glc in GYG1 (19). The loss of the DXD motif and of the coordinating Lys 232 residue abolished the glycosyltransferase activity of R707. By contrast, the mutation of Tyr 215 had no effect on R707 activity (Figure 2B), confirming the lack of self-glycosylation in the mimivirus glycosyltransferase. Like GYG1, the glycosyltransferase activity of R707 required the divalent cation Mn^{2+}. R707 activity was decreased to 19% in absence of Mn^{2+}, to 20% after substitution of Mn^{2+} with Mg^{2+}, and to 10% after substitution of Mn^{2+} with Ca^{2+} (Supplementary Figure 1A). In GYG1 the substitution of Mn^{2+} with Mg^{2+} and Ca^{2+} decreased glycosyltransferase activity to 5% and 10%, respectively (19).

To determine the substrate specificity of R707, we tested the nucleotide-sugar donors UDP-Glc, UDP-GlcNAc because these monosaccharides are the main carbohydrates of mimivirus glycans (8, 10). We also tested UDP-Gal as an epimer of UDP-Glc and UDP-

Xyl because it is used as alternative nucleotide-sugar donor by GYG1 (17). Similarly, we tested the monosaccharides pNP-α-Glc, pNP-α-Xyl, pNP-α-Gal, and pNP-α-Rha as acceptor substrates. UDP-Glc was the only donor substrate compatible with R707 glucosyltransferase activity (Supplementary Figure 1B). The corresponding glucosyltransferase activity was highest towards the α-linked Glc acceptor pNP-α-Glc, but R707 showed also significant glucosyltransferase activity towards pNP-α-Xyl. By contrast, none of the other monosaccharides tested were valid acceptor substrates for R707 (Supplementary Figure1 C). As a comparison, GYG1 can also use UDP-Xyl as donor substrate, although the transfer of Xyl blocks the elongation of the Glc polymer by GYG1 (17).

To characterize the product of R707 catalysis we first subjected 200 pmol of purified reaction product to mass spectrometry analysis. In addition to the acceptor substrate pNP-α-Glc, the elongation products pNP-α-Glc_2 and pNP-α-Glc_3 were detected with a ppm error < 50 ppm (Figure 3), confirming the polymerizing activity of R707. A fragment spectrum of pNP-α-Glc_2 confirmed the presence of the product. To analyze the type of glycosidic linkage catalyzed by R707, reaction products were purified by reverse-phase chromatography allowing a qualitative separation of the main products of the glycosyltransferase reaction. Two main chromatography peaks plus one additional peak recovered were subjected to NMR analysis (Figure 4). All other analyzed peaks showed a too high degree of structural heterogeneity to enable NMR analysis and linkage determination. The glycosidic linkages were identified from the cross peaks in the ROESY spectra. The rotating-frame Overhauser enhancements between the anomeric proton 1' and the protons 6a/6b established the 1,6-linkage for the two main reaction products analyzed (Figure 5, Supplementary Figure 2). Differentiation between the α- and β-configuration at the anomeric centers was based on the $^3J_{1,2}$ coupling constants. Except for the anomeric proton 1' of sample A these couplings could be directly determined from the splitting of the H1 (H1') doublets in the 1D proton

spectrum. For the anomeric proton 1' of sample A, which was obscured by the HOD signal, this coupling was identified from the H2'/H1' cross peak patterns in the DQF-COSY and then extracted from the H2' multiplet in the 1D proton spectrum. For comparison all the H2/H1 (H2'/H1') COSY cross peaks are shown (Figure 5). The neighboring coupling constant between anomeric protons indicates their relative orientation in pyranose structures. In a β configuration a coupling constant of 7-8 Hz is observed, in an α configuration the coupling constant is around 4 Hz (23). The following coupling constants were obtained: sample A: $^3J_{1,2}$ =3.7 Hz; $^3J_{1',2'}$ =3.7 Hz; sample B: $^3J_{1,2}$=3.6 Hz; $^3J_{1',2'}$ =8.0 Hz. The ^{3}J coupling constants confirmed the α-configuration for product A and the β-configuration for product B (Figure 5). The additional reaction product C of R707 was also subjected to NMR analysis; a comparison of the HSQC data with the authentic standard indicated its identity as pNP-α-Glc(α1-4)Glc (Supplementary Figure 3). The NMR analysis thus identified pNP-α-Glc(α1-6)Glc, pNP-α-Glc(β1-6)Glc, and pNP-α-Glc(α1-4)Glc as R707 reaction products. Peaks A, B, and C were further analyzed by GC-MS and negative ion ESI-MS/MS for linkage determination of the pNP-disaccharides. The elution positions in the GC-MS analysis and the fragment ion products of the negative-ion ESI-MS identified a Glc(1-6)Glc linkage for peaks A and B and Glc(1-4)Glc linkage for peak C, corroborating the assignments based on NMR analyses (Figure 6, Supplementary Figure 4).

Mimivirus glycans are mainly found on the heavily glycosylated surface fibers (2, 8). Any additional role for mimivirus glycans in a later stage of infection, however, cannot be ruled out. Considering the formation of linear Glc polymers bearing various linkages by R707, we investigated the functional significance of such Glc polymers and other polysaccharides in the context of mimivirus infection. Upon infection of amoeba by mimivirus, we detected virus factories in 77% of amoeba 4.5 h post infection (Figure 7A). The addition of the α1-4 linked Glc polymer amylose to amoeba decreased mimivirus infection in a dose dependent manner, indicating that α1-4 linked Glc chains

are indeed required for the uptake of the virus by amoeba (Figure 7B). A similar inhibition of virus infection was observed by addition of β1-4 linked GlcNAc in the form of chitin hydrolysate. The localization of the major proportion of GlcNAc is unknown to date as only a small proportion of O-glycan structures contain GlcNAc even though GlcNAc is the major monosaccharide in mimivirus (10). By contrast, addition of the β1-6 linked Glc polymer dextran and of the α1-4 linked mannose polymer mannan had no effect on mimivirus infection (Figure 7B). The competitive inhibition of mimivirus infection by amylose and chitin hydrolysate confirmed the importance of polysaccharides in the capture of mimivirus by amoeba.

DISCUSSION

We have shown that mimivirus R707 is a functional glycosyltransferase transferring Glc from UDP-Glc to Glc through multiple glycosidic linkages. The R707 enzyme has strict donor and acceptor substrate specificities and its glucosyltransferase activity depends on Mn^{2+}. The characterized products of R707 reaction were pNP-α-Glc(α1-6)Glc, pNP-α-Glc(β1-6)Glc, and pNP-α-Glc(α1-4)Glc. Infection of *Acanthamoeba* with mimivirus in presence of soluble amylose and chitin hydrolysate reduced the number of infected amoeba indicating a role for glycan-binding protein interactions in the infection of amoeba.

This is the first report of a single-domain glycosyltransferase capable of forming both α- and β-glycosidic linkages. With few exceptions glycosyltransferases exhibit strict linkage specificities. In some unique cases glycosyltransferases exhibit flexibility regarding the formed glycosidic bond. Human fucosyltransferase III attaches fucose in either α1-3 or α1-4 linkage (24). The α-dystroglycan modifying enzyme LARGE is a bi-functional glycosyltransferase with both xylosyltransferase and glucuronyltransferase activities. LARGE generates repeats of the disaccharide unit (β1-3)Xyl(α1-3)glucuronic acid. In contrast to R707, LARGE consists of two separate glycosyltransferase domains

with each domain responsible for a specific activity (25). We found 58% identity in amino acid residues in the catalytic pocket of GYG1 and R707. The remaining different residues might play a role in directing the inverting and retaining mechanism of linkage formation. In inverting glycosyltransferases an active-site side chain is needed as base catalyst for deprotonation the incoming nucleophile of the acceptor substrate. In retaining glycosyltransferases a nucleophile within the active site is needed to allow a double-displacement reaction (26). Crystallization of R707 and further studies of the catalytic pocket in presence and absence of the acceptor substrate may identify the residues possibly serving as base catalyst and nucleophilic side chains in the transfer of Glc. This could explain the flexibility of R707 in mediating distinct linkages and configurations.

GYG1 catalysis produces linear α1-4 Glc polymers. Branching occurs every 8-10 residues and is catalyzed by the glycogen branching enzyme GBE1, which catalyzes the α1-6 addition of Glc (27). We identified pNP-α-Glc(α1-6)Glc and pNP-α-Glc(β1-6)Glc and pNP-α-Glc(α1-4)Glc as products of R707 catalysis. The Glc trimers we identified were linear chains of Glc. This goes in hand with data published on mimivirus O-glycans, where all identified hexose polymers were linear (10). Glycogen is the universal molecule for Glc storage. Glycogen synthesis has been reported for a large number of species, from bacteria to yeast to mammals (17, 28, 29). In mammals, glycogen synthesis is initiated by self-glucosylation of GYG1 in the cytosol (27, 29). In contrast to GYG1 mimivirus R707 did not exhibit self-glucosylating activities and a disruption of the Tyr215 residue, corresponding to Tyr 195 in GYG1, had no influence on R707 activity. Mimivirus depends on energy supply from amoeba and as the virus needs immediate energy availability energy storage in form of glycogen hardly makes sense (30). It is more likely that GYG1 and R707 showed divergent evolution where GYG1 evolved to being the core protein of glycogen and R707 to being a glycosyltransferase involved in decorating other mimivirus proteins with Glc polymers.

An independent glycosylation machinery allows mimivirus to form virus-specific glycan structures. These specific glycans are probably needed for virus-host interactions and mimivirus infection. Mimivirus polysaccharides could mimic the cell wall of gram-positive bacteria and thereby increase the chance of phagocytosis by *Acanthamoeba* (31-33). *Acanthamoeba spp.* express various lectins of which some are known to be involved in phagocytosis of their prey (34-36). Parallel addition of amylose or chitin hydrolysate during mimivirus infection of amoeba reduces infection in a dose dependent manner. This effect is not seen in presence of dextran, mannan or poly vinyl alcohol suggesting specific receptor occupancy on the surface of amoeba or mimivirus resulting in reduced mimivirus uptake. In addition, flexible and diverse surface glycosylation might allow mimivirus to target different host species, as observed for *Legionella pneumophila*. *Legionella pneumophila* is taken up by both *Acanthamoeba castellanii* and *Naegleria lovaniensis* via receptor-mediated endocytosis but through binding to the different receptors, mannose-binding receptor on the one hand, and N-acetyl-D-galactosamine receptor on the other hand (36).

The results described in this study present a basis for future investigations for a better understanding of the mechanism of glycosidic bond formation in glycosyltransferases. Our results further establish the existence of a host-independent glycosylation machinery in mimivirus. Furthermore, the findings presented underline the importance of glycan-protein interactions in mimivirus infection.

ACKNOWLEDGMENTS

We thank Luca Plan for his support with the mutagenesis experiments.

DECLARATION OF INTEREST

The authors declare no conflict of interest with the content of this article.

FUNDING INFORMATION

This work was supported by the Research Credit of the University of Zurich to AR and by the Swiss National Foundation grant 310030_149949 to TH.

AUTHOR CONTRIBUTION STATEMENT

Anna Rommel, Andreas Hüslmeier, and Thierry Hennet designed the experiments. Anna Rommel and Thierry Hennet wrote the manuscript with contributions from Andreas Hülsmeier and Simon Jurt. Andreas Hüslmeier carried out all MS analyses. Simon Jurt did the NMR experiments. All other experiments were done by Anna Rommel.

REFERENCES

1. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, et al. A giant virus in amoebae. Science. 2003;299(5615):2033. Epub 2003/03/29.

2. Kuznetsov YG, Xiao C, Sun S, Raoult D, Rossmann M, McPherson A. Atomic force microscopy investigation of the giant mimivirus. Virology. 2010;404(1):127-37. Epub 2010/06/17.

3. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. Science. 2004;306(5700):1344-50. Epub 2004/10/16.

4. Vigerust DJ, Shepherd VL. Virus glycosylation: role in virulence and immune interactions. Trends in microbiology. 2007;15(5):211-8. Epub 2007/04/03.

5. Van Etten JL, Gurnon JR, Yanai-Balser GM, Dunigan DD, Graves MV. Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the endoplasmic reticulum and Golgi. Biochimica et biophysica acta. 2010;1800(2):152-9. Epub 2009/08/06.

6. De Castro C, Molinaro A, Piacente F, Gurnon JR, Sturiale L, Palmigiano A, et al. Structure of N-linked oligosaccharides attached to chlorovirus PBCV-1 major capsid protein reveals unusual class of complex N-glycans. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(34):13956-60. Epub 2013/08/07.

7. Parakkottil Chothi M, Duncan GA, Armirotti A, Abergel C, Gurnon JR, Van Etten JL, et al. Identification of an L-rhamnose synthetic pathway in two nucleocytoplasmic large DNA viruses. Journal of virology. 2010;84(17):8829-38. Epub 2010/06/12.

8. Piacente F, Marin M, Molinaro A, De Castro C, Seltzer V, Salis A, et al. The giant DNA virus Mimivirus encodes a pathway for the biosynthesis of the unusual sugar 4-amino-4,6-dideoxy-D-glucose (viosamine). The Journal of biological chemistry. 2011. Epub 2011/12/14.

9. Piacente F, Bernardi C, Marin M, Blanc G, Abergel C, Tonetti MG. Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNA virus Mimivirus. Glycobiology. 2014;24(1):51-61. Epub 2013/10/11.

10. Hulsmeier AJ, Hennet T. O-Linked glycosylation in Acanthamoeba polyphaga mimivirus. Glycobiology. 2014;24(8):703-14. Epub 2014/05/06.

11. Luther KB, Hulsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennet T. Mimivirus collagen is modified by a bifunctional lysyl hydroxylase and glycosyltransferase enzyme. The Journal of biological chemistry. 2011. Epub 2011/11/03.

12. Braman J, Papworth C, Greener A. Site-directed mutagenesis using double-stranded plasmid DNA templates. Methods Mol Biol. 1996;57:31-44. Epub 1996/01/01.

13. Hulsmeier AJ, Welti M, Hennet T. Glycoprotein maturation and the UPR. Methods in enzymology. 2011;491:163-82. Epub 2011/02/19.

14. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, et al. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. Journal of biomolecular NMR. 1995;6(2):135-40. Epub 1995/09/01.

15. Palma AS, Liu Y, Zhang H, Zhang Y, McCleary BV, Yu G, et al. Unravelling glucan recognition systems by glycome microarrays using the designer approach and mass spectrometry. Mol Cell Proteomics. 2015;14(4):974-88.

16. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nature methods. 2012;9(7):676-82. Epub 2012/06/30.

17. Lomako J, Lomako WM, Whelan WJ. Glycogenin: the primer for mammalian and yeast glycogen synthesis. Biochimica et biophysica acta. 2004;1673(1-2):45-55. Epub 2004/07/09.

18. Ramakrishnan B, Boeggeman E, Ramasamy V, Qasba PK. Structure and catalytic cycle of beta-1,4-galactosyltransferase. Current opinion in structural biology. 2004;14(5):593-600. Epub 2004/10/07.

19. Gibbons BJ, Roach PJ, Hurley TD. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. Journal of molecular biology. 2002;319(2):463-77. Epub 2002/06/08.

20. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. Nature protocols. 2009;4(1):1-13. Epub 2009/01/10.

21. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols. 2015;10(6):845-58. Epub 2015/05/08.

22. Legendre M, Santini S, Rico A, Abergel C, Claverie JM. Breaking the 1000-gene barrier for Mimivirus using ultra-deep genome and transcriptome sequencing. Virology journal. 2011;8:99. Epub 2011/03/08.

23. Duus J, Gotfredsen CH, Bock K. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. Chem Rev. 2000;100(12):4589-614. Epub 2001/12/26.

24. Rini J, Esko J, Varki A. Glycosyltransferases and Glycan-processing Enzymes. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

25. Inamori K, Yoshida-Moriguchi T, Hara Y, Anderson ME, Yu L, Campbell KP. Dystroglycan function requires xylosyl- and glucuronyltransferase activities of LARGE. Science. 2012;335(6064):93-6. Epub 2012/01/10.

26. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. Annual review of biochemistry. 2008;77:521-55. Epub 2008/06/04.

27. Smythe C, Cohen P. The discovery of glycogenin and the priming mechanism for glycogen biogenesis. European journal of biochemistry / FEBS. 1991;200(3):625-31. Epub 1991/09/15.

28. Okita TW, Rodriguez RL, Preiss J. Biosynthesis of bacterial glycogen. Cloning of the glycogen biosynthetic enzyme structural genes of Escherichia coli. The Journal of biological chemistry. 1981;256(13):6944-52. Epub 1981/07/10.

29. Wilson WA, Roach PJ, Montero M, Baroja-Fernandez E, Munoz FJ, Eydallin G, et al. Regulation of glycogen metabolism in yeast and bacteria. FEMS microbiology reviews. 2010;34(6):952-85. Epub 2010/04/24.

30. Mutsafi Y, Fridmann-Sirkis Y, Milrot E, Hevroni L, Minsky A. Infection cycles of large DNA viruses: Emerging themes and underlying questions. Virology. 2014. Epub 2014/07/06.

31. Klose T, Herbst DA, Zhu H, Max JP, Kenttamaa HI, Rossmann MG. A Mimivirus Enzyme that Participates in Viral Entry. Structure. 2015;23(6):1058-65. Epub 2015/05/20.

32. Sobhy H, Scola BL, Pagnier I, Raoult D, Colson P. Identification of giant Mimivirus protein functions using RNA interference. Frontiers in microbiology. 2015;6:345. Epub 2015/05/15.

33. Rodrigues RA, Dos Santos Silva LK, Dornas FP, de Oliveira DB, Magalhaes TF, Santos DA, et al. Mimivirus Fibrils Are Important for Viral Attachment to the Microbial World by a Diverse Glycoside Interaction Repertoire. Journal of virology. 2015;89(23):11812-9. Epub 2015/09/18.

34. Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, et al. Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome biology. 2013;14(2):R11. Epub 2013/02/05.

35. Medina G, Flores-Martin S, Fonseca B, Otth C, Fernandez H. Mechanisms associated with phagocytosis of Arcobacter butzleri by Acanthamoeba castellanii. Parasitology research. 2014;113(5):1933-42. Epub 2014/03/22.

36. Declerck P, Behets J, De Keersmaecker B, Ollevier F. Receptor-mediated uptake of Legionella pneumophila by Acanthamoeba castellanii and Naegleria lovaniensis. Journal of applied microbiology. 2007;103(6):2697-703. Epub 2007/09/14.

37. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990;215(3):403-10. Epub 1990/10/05.

38. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic acids research. 2015;43(Database issue):D222-6. Epub 2014/11/22.

39. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology. 2011;7:539. Epub 2011/10/13.

FIGURE LEGENDS

FIGURE 1. Sequence homology between mimivirus R707 and eukaryotic glycogenin-1 proteins. (A) Alignment of R707 and glycogenin-1 (GYG1) from different species using protein BLAST and the non-redundant protein sequences database (37, 38). The multiple sequence alignment of R707 and GYG1 homologs was generated using Clustal Omega (39). Identical residues are shaded in black (≥ 4 species) or in grey (<4 species). The DXD motif and catalytic tyrosine of GYG1 are marked with dashed boxes. Mutated residues of R707 are marked with open triangles. (B) Catalytic center of mouse Gyg1 coordinating UDP-Glc and Mn^{2+} (adapted from (19)). Residues conserved in R707 are marked in blue. (C) Schematic representation of full-length R707 (281 amino acids long) and of its predicted glycosyltransferase domain spanning residues 4 to 265.

FIGURE 2. Glycosyltransferase activity of R707. (A) Glucosyltransferase activity of recombinant R707 on pNP-α-Glc, pNP-β-Glc and self-glucosylation of R707 in absence of an acceptor substrate. (B) Glucosyltransferase activity of wildtype R707 in comparison to R707 including the amino acid substitutions D103S, Y215F, and K232R. Two independent experiments with triplicates in each instance were done. Results are shown as mean ± SEM.

FIGURE 3. (A) MALDI-MS analysis of R707 assay products. 200 pmol of glycosyltransferase assay product were subjected to MALDI-MS analysis. The molecular ions of assay products are labeled with their glycan annotation. Glucose residues are represented by open circle. Ppm error (pNP-Glc_2) = 34 ppm, ppm error (pNP-Glc_3) = 15 ppm. (B) MS2 spectrum of m/z 486.1. The inserted symbols represent the proposed assignments.

FIGURE 4. R707 product purification. Approximately 300 nmol of R707 assay products were separated on a Hypersil C18 column. Peaks A, B, and C were subjected to further structural analysis by NMR.

FIGURE 5. Linkage analysis of R707 assay product by NMR with (A) referring to peak A, figure 4 and (B) referring to peak B from figure 4. Two unidentified product species were subjected to proton NMR TOCSY (middle panel, top) for identification of the two spin systems from the sugar units. Configuration between pNP and Glc (COSY, lower panel) and configuration between Glc and Glc (COSY, upper panel) by expansion of the DQF-COSY cross peaks between the anomeric proton and proton at position 2 or 2' respectively allowing extraction of the ^{3}J coupling constants and discrimination between the alpha ($^{3}J_{(1,2)}$~3 Hz) and beta ($^{3}J_{(1,2)}$~7 Hz) configuration. Rotating-frame Overhauser enhancement between aromatic proton A and anomeric proton 1 (ROESY, left panel) for linkage determination between pNP and Glc. Rotating-frame Overhauser enhancement between anomeric proton 1' and protons 6a/6b (ROESY, middle panel, bottom) for linkage determination between Glc and Glc. The experiment was done twice.

FIGURE 6. Product linkage analysis of peaks A, B, and C using GC-MS. Partially methylated alditol acetates (PMMA) were prepared from the R707 reaction products and resolved by GC-MS. Extracted chromatograms at m/z 118.061. Characteristic hexoses are shown. The elution positons of 2-, 3-, 4-, and 6-O-substituted PMMA were determined by comparison with PMMAs prepared from authentic reference compounds.

FIGURE 7. Mimivirus uptake competition assay. (A) Amoeba were infected at a multiplicity of 250, (B) and in combination with 1 μg, 5 μg, or 10 μg of polysaccharides or poly vinyl alcohol (PVA). Virus factories per total number of amoeba per viewfield were quantified 4.5 h post infection using an Axiovert 200M microscope and Fiji cell counter. Two independent experiments quantifying five viewfields per condition were done. Results are shown as mean ± SEM.

SUPPLEMENTARY FIGURE 1. Co-factor and substrate specificity of R707. (A) Assays with Mn^{2+}, Mg^{2+}, or Ca^{2+} and in absence of divalent metal ions. (B) Nucleotide sugar

donor specificity using UDP-Glc or UDP-GlcNAc or UDP-Gal or UDP-Xyl. (C) Acceptor substrate specificity using pNP-α-Glc, pNP-α-Xyl, pNP-α-Gal, or pNP-α-Rha. Two independent experiments with triplicates in each instance were performed. Results are shown as mean ± SEM.

SUPPLEMENTARY FIGURE 2 Heteronuclear single quantum coherence (HSQC) signals of peaks A and B

SUPPLEMENTRAY FIGURE 3. Heteronuclear single quantum coherence (HSQC) signals of peak C and pNP-maltose.

SUPPLEMENTARY FIGURE 4. Product analysis of peaks A, B, and C by negative-ion ESI-MS. 60 pmol of products were subjected to analysis. The inserted symbols represent the proposed assignments.

TABLE I PCR primers used to construct mutated R707. Underlined bases mark the introduced mutations. F: forward primer, R: reverse primer.

Mutated Residue	Primer sequence
mR707-Asp103Ser	F 5′-CGATAAAATTATTTTATTAGATTTAAGCATGATAATTGC-3′
	R 5′-GCAATTATCATGCTTAAATCTAATAAAATAATTTTATCG-3′
mR707-Tyr215Phe	F 5′-GGGTTGACACATCGAGTAAAAAAGTTCCATTATACAATTGA-3′
	R 5′-CAATTGTATAATGGAACTTTTTTACTCGATGTGTCAACCC-3′
mR707-Lys232Arg	F 5′-CATTTTTCTAGTTCCTATCGACCATGGAACAGATTAAATTCCGAC-3′
	R 5′-CGGAATTTAATCTGTTCCATGGTCGATAGGAACTAGAAAAATGA-3′

Figure 1

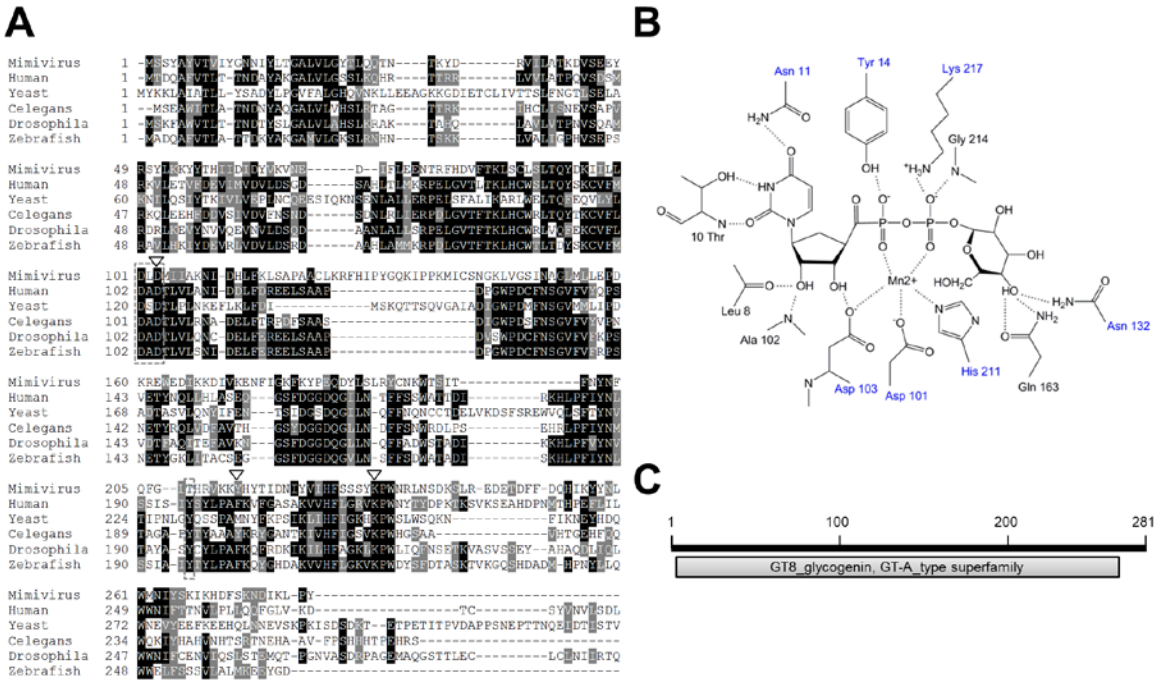

A
Mimivirus
Human
Yeast
Celegans
Drosophila
Zebrafish
B
Asn 11
Tyr 14
Lys 217
Gly 214
10 Thr
Leu 8
Ala 102
Mn2+
Asp 103
Asp 101
His 211
Gln 163
Asn 132
C
1
100
200
281
GT8_glycogenin, GT-A_type superfamily

Figure 2

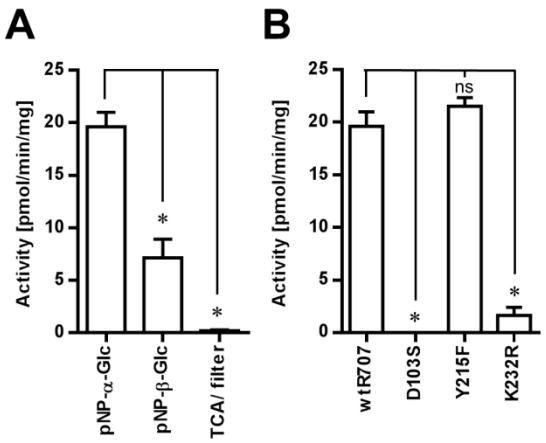

A
B
Activity [pmol/min/mg]
25
20
15
10
5
0
*
pNP-α-Glc
pNP-β-Glc
TCA/ filter
ns
wtR707
D103S
Y215F
K232R

Figure 3

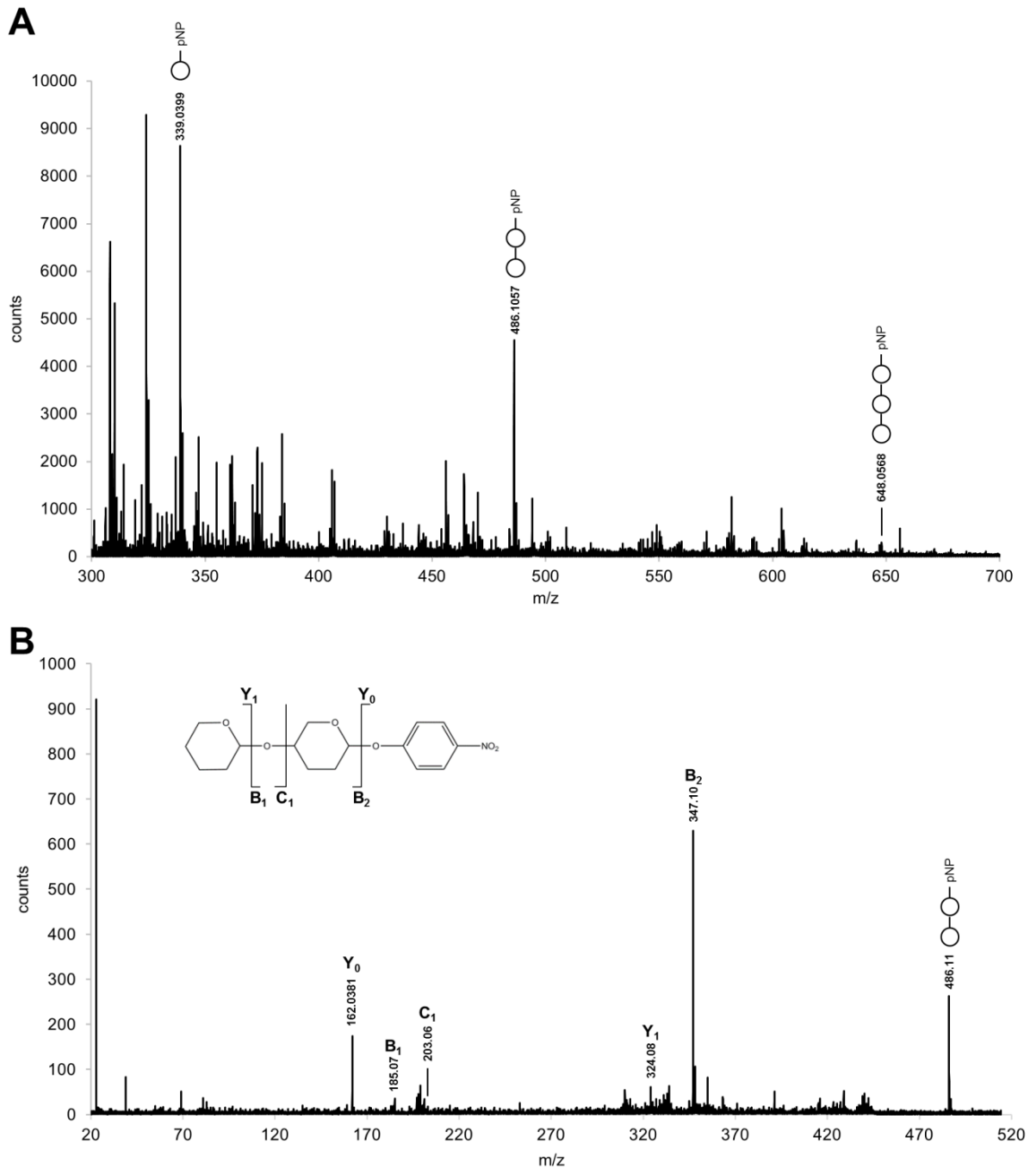

A
pNP
339.0399
pNP
486.1057
pNP
648.0568
counts
m/z
B
Y1
Y0
B1
C1
B2
NO2
B2
347.10
pNP
486.11
Y0
162.0381
B1
185.07
C1
203.06
Y1
324.08
counts
m/z

Figure 4

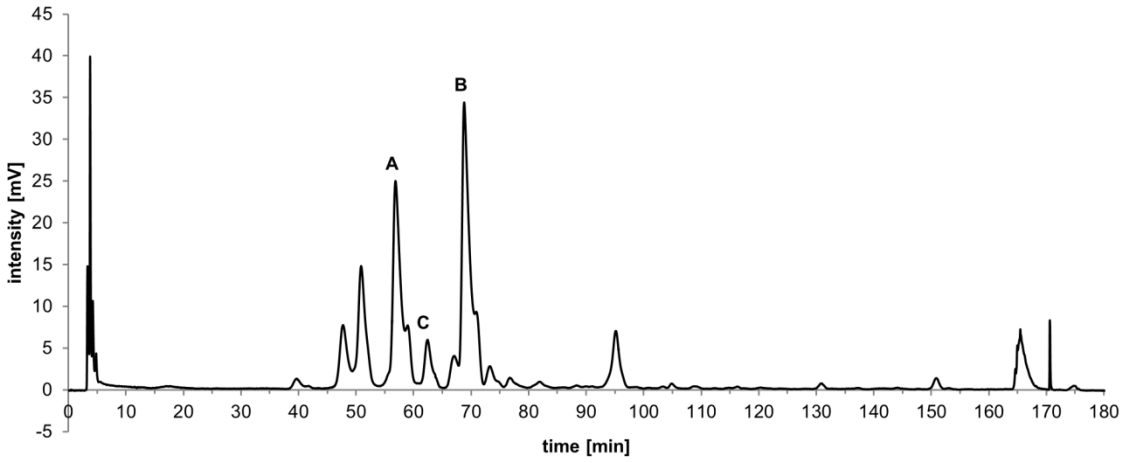

intensity [mV]
time [min]
A
B
C
45
40
35
30
25
20
15
10
5
0
-5
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180

Figure 5

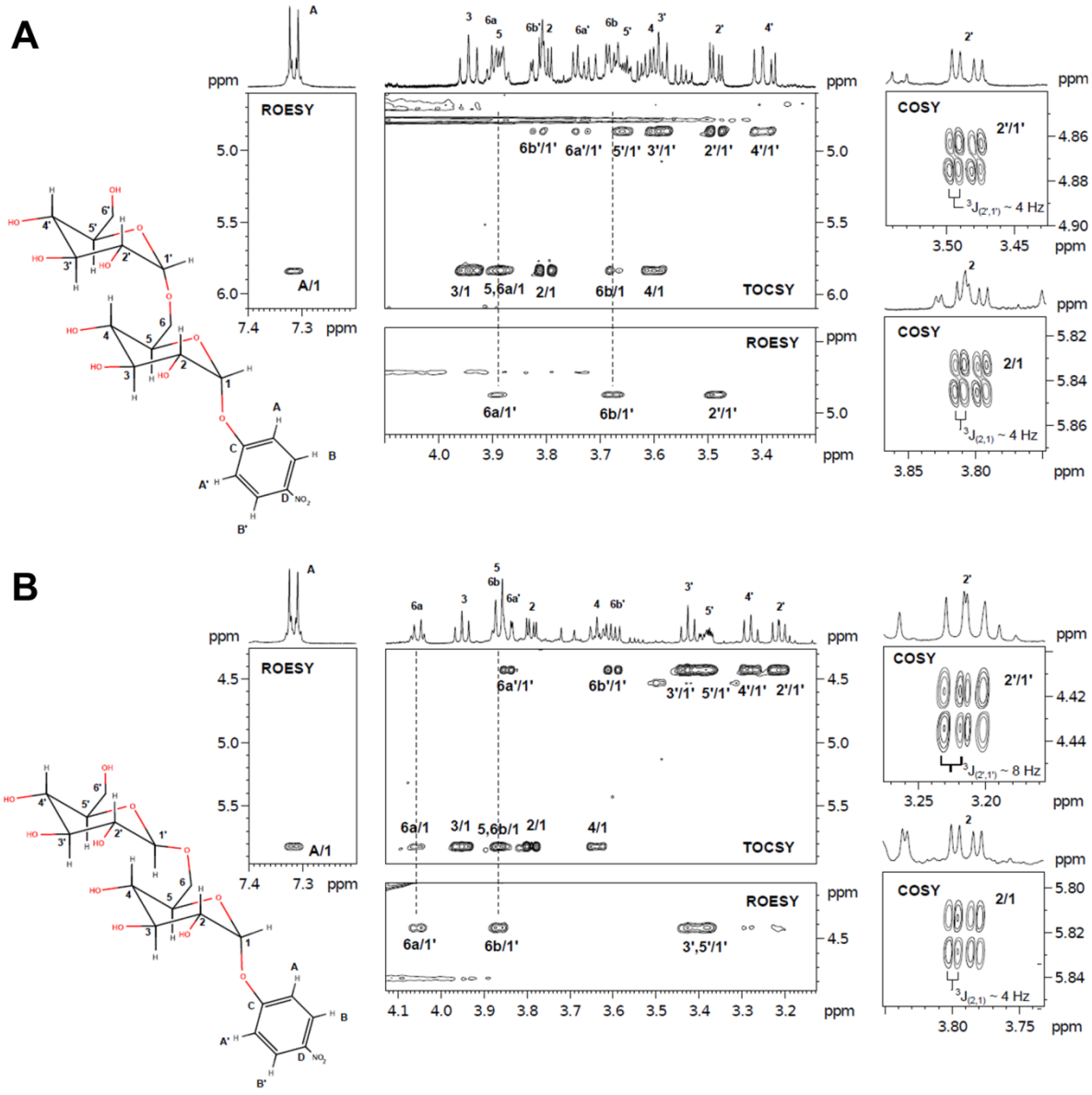

A
ROESY
A/1
TOCSY
6b'/1'
6a'/1'
5'/1'
3'/1'
2'/1'
4'/1'
3/1
5,6a/1
2/1
6b/1
4/1
6a/1'
6b/1'
COSY
2'/1'
$^3J_{(2',1')}$ ~ 4 Hz
2/1
$^3J_{(2,1)}$ ~ 4 Hz
B
ROESY
A/1
TOCSY
6a'/1'
6b'/1'
3'/1'
5'/1'
4'/1'
2'/1'
6a/1
3/1
5,6b/1
2/1
4/1
6a/1'
6b/1'
3',5'/1'
COSY
2'/1'
$^3J_{(2',1')}$ ~ 8 Hz
2/1
$^3J_{(2,1)}$ ~ 4 Hz

Figure 6

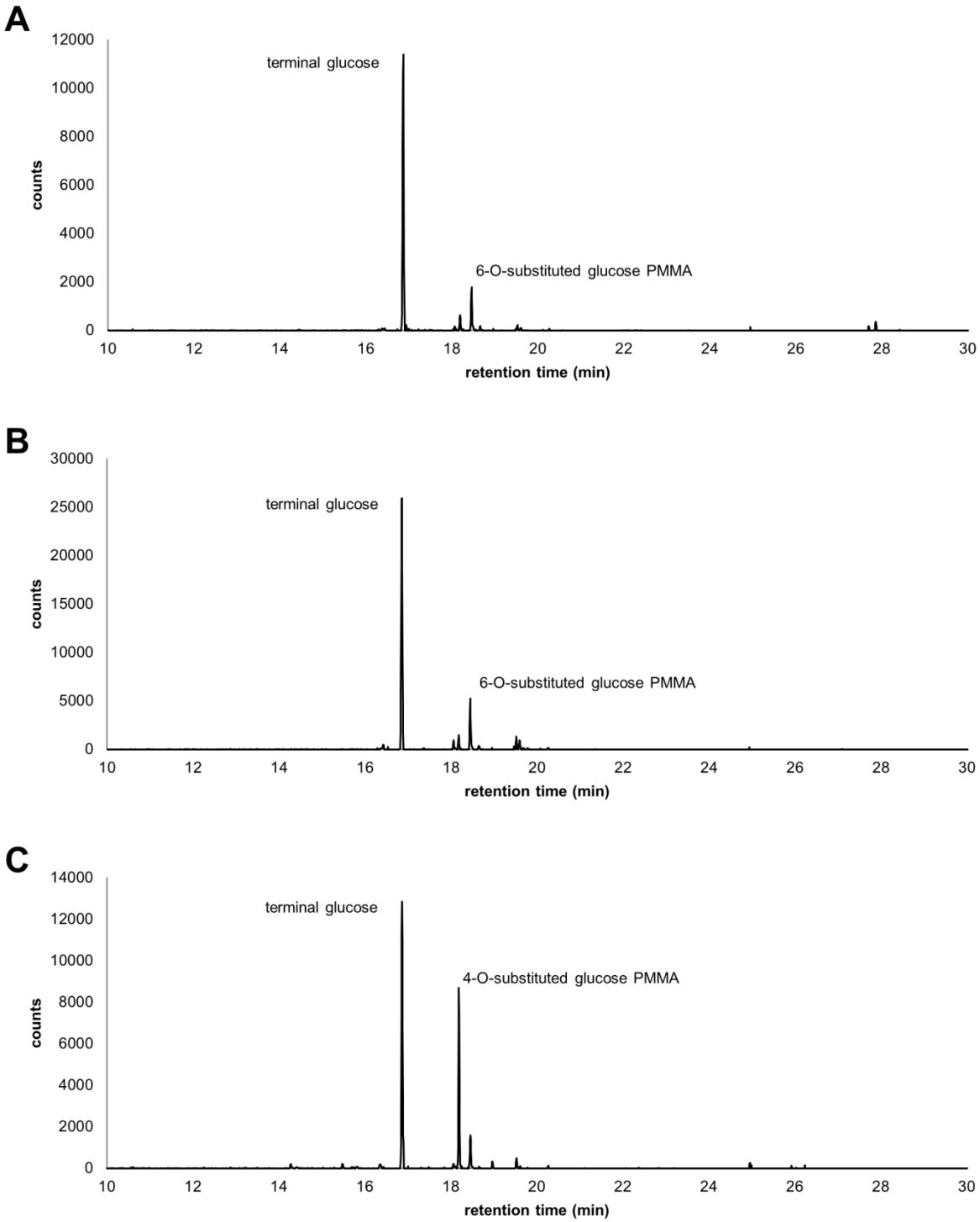
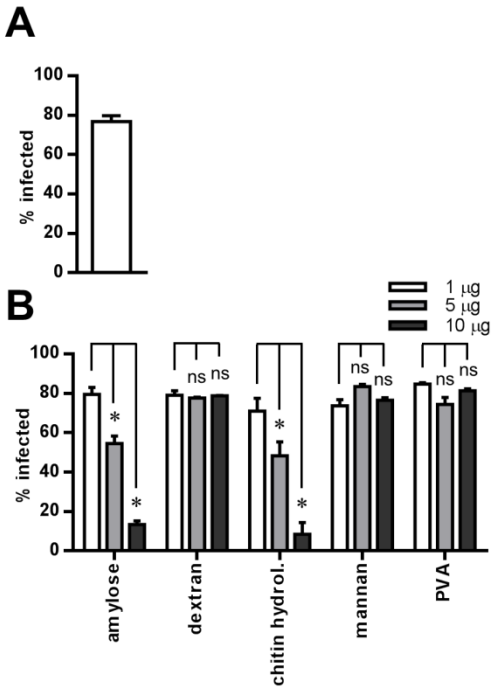
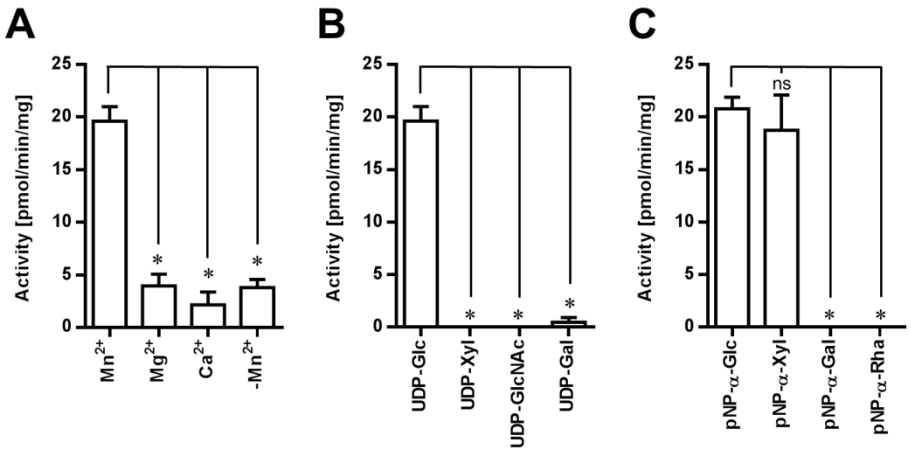

A
terminal glucose
6-O-substituted glucose PMMA
counts
retention time (min)
B
terminal glucose
6-O-substituted glucose PMMA
counts
retention time (min)
C
terminal glucose
4-O-substituted glucose PMMA
counts
retention time (min)

Figure 7

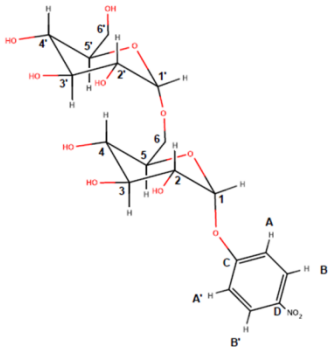

A
100
80
60
40
20
0
% infected
B
1 μg
5 μg
10 μg
100
80
60
40
20
0
% infected
*
*
ns ns
*
*
ns
ns
ns
ns
amylose
dextran
chitin hydrol.
mannan
PVA

Supplementary Figure 1



Supplementary Figure 2

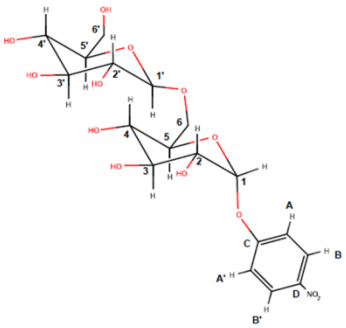
A



Position	1	2	3	4	5	6a	6b	1’	2’
δ(^{1}H) / ppm	5.84	3.80	3.95	3.60	3.88	3.90	3.68	4.87	3.48
δ(^{13}C) / ppm	99.1	73.4	75.7	71.9	74.1	68.0	68.0	100.2	73.9

Position	3’	4’	5’	6a’	6b’	A,A’	B,B’	C	D
δ(^{1}H) / ppm	3.59	3.40	3.66	3.74	3.82	7.31	8.28		
δ(^{13}C) / ppm	75.6	72.0	74.3	63.0	63.0	119.4	128.6	164.0	145.1

B



Position	1	2	3	4	5	6a	6b	1’	2’
δ(^{1}H) / ppm	5.82	3.79	3.95	3.64	3.86	4.06	3.87	4.42	3.21
δ(^{13}C) / ppm	99.2	73.4	75.4	71.5	74.6	70.7	70.7	105.2	75.7

Position	3’	4’	5’	6a’	6b’	A,A’	B,B’	C	D
δ(^{1}H) / ppm	3.43	3.28	3.39	3.84	3.60	7.32	8.28		
δ(^{13}C) / ppm	78.3	72.2	78.5	63.4	63.4	119.5	128.6	164.0	145.1

Supplementary Figure 3

HSQC signals Peak C

δ(^{1}H) / ppm	3.75	3.74	3.82	3.4	3.8	3.78	3.58	3.68	3.66
δ(^{13}C) / ppm	62.9	63.1	63.1	72.0	73.5	74.1	74.4	75.4	75.5

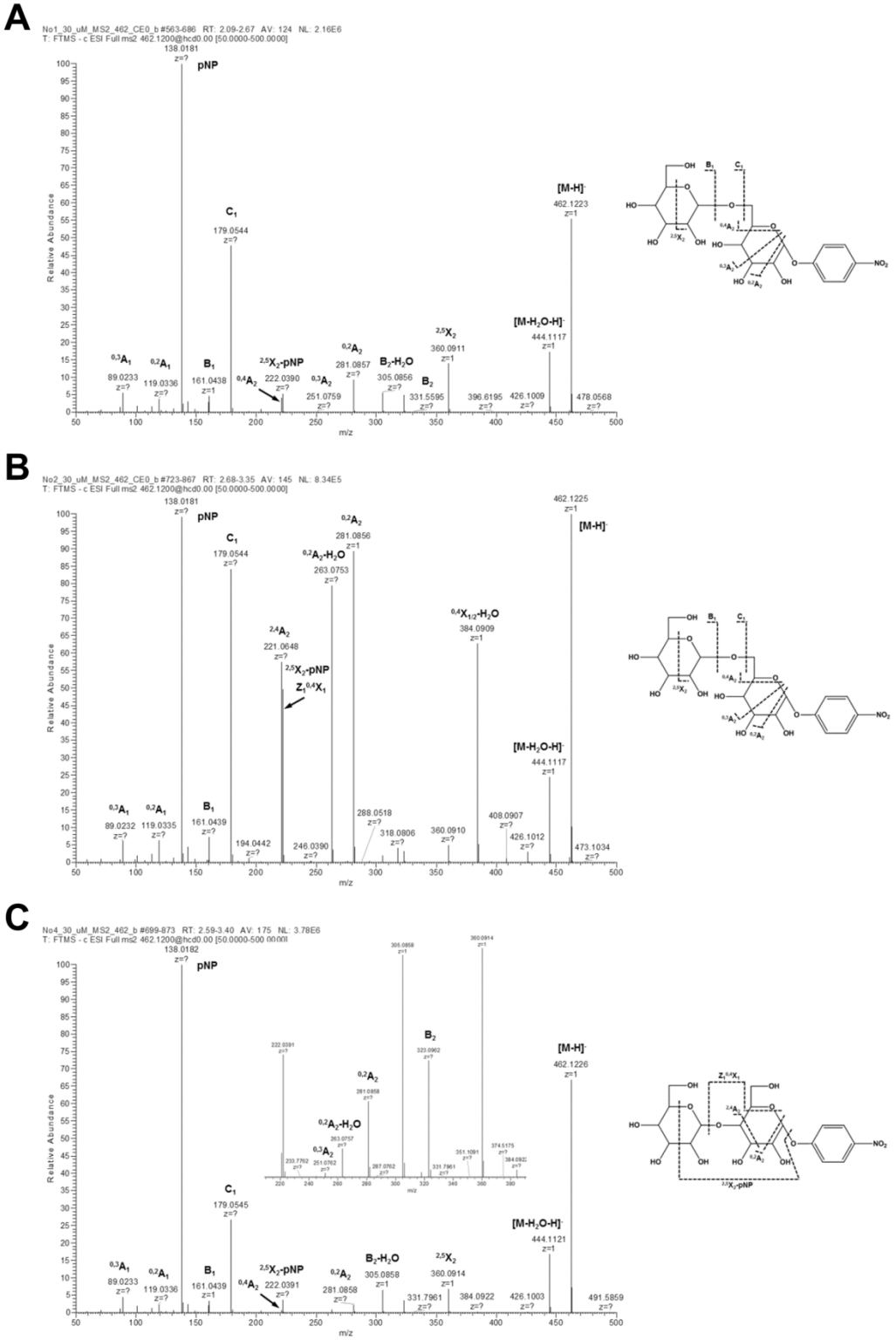
δ(^{1}H) / ppm	4.21	3.76	5.8	5.42	7.28	8.23
δ(^{13}C) / ppm	76	79.1	99.1	102.4	119.4	128.7

HSQC signals pNP-maltose

δ(^{1}H) / ppm	3.76	3.74	3.83	3.41	3.81	3.8	3.57	3.68	3.67
δ(^{13}C) / ppm	62.9	63.2	63.2	72.0	73.5	74.1	74.4	75.4	75.5

δ(^{1}H) / ppm	4.22	3.77	5.82	5.43	7.31	8.27
δ(^{13}C) / ppm	76.1	79	99.1	102.4	119.5	128.7

Supplementary Figure 4



MANUSCRIPT II: A UDP-GLCNAC HYDROLYZING ENZYME IN MIMIVIRUS

***Acanthamoeba polyphaga* mimivirus encodes a UDP-GlcNAc hydrolyzing enzyme with glycosyltransferase characteristics**

Manuscript in preparation

Short title: A UDP-GlcNAc hydrolyzing enzyme in mimivirus

Keywords: giant virus / glycosylation/ hydrolysis/ glycosyltransferase

Anna Rommel[1], Thierry Hennet[1*]

[1]Institute of Physiology, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

*Corresponding author

E-mail: thennet@access.uzh.ch

Experiments were designed by AR and TH. Experiments were done by AR. The manuscript was written by AR.

ABSTRACT

The giant *Acanthamoeba polyphaga* mimivirus encodes 1'262 genes. Several of these genes are part of a thought to be virus-specific glycosylation machinery, like enzymes involved in the synthesis of nucleotide-diphospho sugars and putative glycosyltransferases. Among them we identified L373 as putative glycosyltransferase with a predicted mannosyl- and N-acetylglucosaminyltransferase activity. We could show specific hydrolysis of UDP-GlcNAc, but not GDP-Man, in a divalent metal ion dependent manner. Taking into account that mimivirus expresses a functional pathway for the synthesis of UDP-GlcNAc our data gives further evidence for the presence of a host-independent UDP-GlcNAc utilization of mimivirus.

ABBREVIATIONS

Fuc	fucose
Gal	galactose
Glc	glucose
GlcNAc	N-acetylglucosamine
HexNAc	N-acetylhexosamine
Man	mannose
pNP	4-nitrophenyl
Rha	rhamnose
UDP	uridine diphosphate
Vio	viosamine
Xyl	xylose

INTRODUCTION

Acanthamoeba polyphaga mimivirus is a giant virus with a 1.2 Mbp genome encoding 1'262 genes (1). The icosahedral particle with a diameter of 400 nm is covered with glycosylated protein fibers (2). Many mimivirus genes code for factors typical for cellular life and are novel in viruses (1).

Among these genes several are involved in the synthesis of nucleotide-diphospho sugars, namely UDP-N-acetylglucosamine (GlcNAc), UDP-rhamnose (Rha), and UDP-viosamine (Vio) (3-5). Three enzymes are involved in the synthesis of UDP-GlcNAc: a glutamine-fructose-6-phosphate transaminase, a glucosamine-6-phosphate N-acetyltransferase, and a UDP-GlcNAc pyrophosphorylase. The viral pathway follows both an eukaryotic-like strategy and a prokaryotic-like strategy (3). The nucleotide sugars synthesized host-independently reflect the monosaccharide composition of mimivirus: GlcNAc, Rha, Vio, together with glucose (Glc), are the major monosaccharides in mimivirus. Small amounts of fucose (Fuc) and xylose (Xyl) and methylated variations thereof are also found (5, 6). Traces of ribose, mannose (Man), arabinose and galactose (Gal) are also detected in mimivirus (5). Although GlcNAc is highly abundant in mimivirus in a study analyzing mimivirus O-glycan structures GlcNAc is underrepresented. The majority of identified mimivirus O-glycans are linear hexose (probably Glc) polymers with varying reducing end residues. Some branched structures were also identified consisting of a hexose at the reducing end and a terminal HexNAc (probably GlcNAc) residue and a terminal pentose residue linked to HexNAc (6).

Analysis of the mimivirus genome suggests that in addition to the nucleotide-diphospho sugar synthesizing machinery mimivirus encodes at least 11 putative glycosyltransferases that share homologies with eukaryotic, prokaryotic, and archaeal glycosyltransferases. The first characterized mimivirus glycosyltransferase is encoded

by ORF L230. L230 is a bifunctional enzyme hydroxylating lysine and glucosylating hydroxylysine in mimivirus collagen substrates (7). Nothing is known about the functionality and role of the other putative mimivirus glycosyltransferases to date. In this study we investigated the functionality of mimivirus ORF L373 which is annotated as putative glycosyltransferase.

EXPERIMENTAL PROCEDURES

Expression and purification of recombinant L373 protein - Recombinant L373 was expressed in *Spodoptera frugiperda* Sf9 cells (Invitrogen, Carlsbad, CA) as FLAG-fusion protein in pFastBac-FLAG(C) vector (Invitrogen). Viral DNA was purified using phenol/ chloroform (Roth, Karlsruhe, Germany). The DNA sequence corresponding to mimivirus L373 (RefSeq ID NC_014649.1) was amplified using the primers: 5'-GCCGGATCCATTCCTAATATTATTCAT-3' and 5'-CGCTCTAGATC-AAGACTTGTTAAAT-3' containing BamHI and XbaI restriction sites, respectively (underlined bases). Recombinant baculoviruses were produced in Sf9 cells as described in Hennet et al. (8).

Control of Protein Expression - For protein separation 5 µl of post nuclear supernatant from *Sf-9* cells overexpressing L373 were subjected to 14% SDS-PAGE. Proteins were transferred to nitrocellulose membrane (GE Healthcare, Glattbrugg, Switzerland) in semi-dry manner. After transfer the blot was blocked in 10% bovine serum albumin (BSA) (Sigma-Aldrich, Buchs, Switzerland) in PBS-Tween (0.01%) (PBS-T) at room temperature for 1 h. Primary anti-FLAG antibody (Sigma-Aldrich) was diluted 1:5'000 in PBS-T/ 1% BSA and incubated at 4°C overnight. After three 5 min wash steps in PBS-T the blot was incubated with secondary peroxidase conjugated goat anti-rabbit IgG (Sigma-Aldrich), diluted 1:10'000 in PBS-T/ 1% BSA at room temperature for 1 h. After three 5 min wash steps in PBS-T the blot was developed with SuperSignal chemiluminescent substrate (Thermo Scientific, Waltham, MA).

Enzymatic activity – Glycosyltransferase activity was measured in 20 mM Tris-HCl pH 7.4 containing 0.5 µl 20 µCi/ml UDP/GDP-[^{14}C]-sugar, 30 µM UDP/GDP-sugar, 10 mM $MnCl_2$ (Sigma-Aldrich), 1 mM DTT (Sigma-Aldrich), 20 mM 4-nitrophenyl (pNP)-sugar in DMSO (Sigma-Aldrich) or 20 µg mimivirus peptides or 20 µg mimivirus proteins, and 10 µg post nuclear supernatant from *Sf-9* cells overexpression L373 or from *Sf-9* cells infected with the empty control vector. The nucleotide-sugars tested as donor substrates were UDP-Glc, UDP-GlcNAc, and GDP-Man. The pNP-sugars tested as acceptor substrates were pNP-α-Fuc, pNP-α-Gal, pNP-α/β-Glc, pNP-α/β-GlcNAc, pNP-α-Rha, pNP-α-Xyl (Sigma-Aldrich). The peptides tested were synthesized from mimivirus ORFs L357, L410, R135, and R459. The proteins tested, ORF L425, R459, and R856, were expressed in Escherichia coli BL21 De3 (Novagen, Nottingham, UK) as His_{10}-fusion proteins and purified by affinity chromatography substantially as described previously (7). Assays were incubated at 30°C for 4 h. Assays without acceptor substrate were stopped by addition of ice cold 900 µl H_2O, and purified over Ag1-X8 resin (Biorad, Hercules, CA). Samples were added to the column, flow-through was collected and product was eluted with 1 ml H_2O. Assays containing pNP-sugar as acceptor were stopped by addition of ice cold 500 µl H_2O, and purified over SepPak-C18 cartridges (Waters, Milford, MA). Samples were added to the cartridges, washed with 10 ml of H_2O, and eluted with 5 ml of methanol (Sigma-Aldrich). Radioactivity was measured after addition of 10 ml of IRGASAFE scintillation liquid (Perkin-Elmer, Waltham, MA) in a Tri-Carb 2900TR scintillation counter (Perkin-Elmer). Assays containing mimivirus peptides or proteins as acceptor were stopped with 500 µl 5% TCA/ 5% phosphotungstic acid and incubated on ice for 30 min. Precipitates were recovered on 1 µm glass filters (VWR, Radnor, PA) using a vacuum manifold and radioactivity was measured as described above.

Statistical analysis - ANOVA with multiple comparisons was used and a P-value of $P < 0.05$ was accepted as statistically significant. GraphPad Prism 6 (GraphPad Software, Inc., La Jolla, CA) was used for statistical analysis. Results are presented as mean ± SEM.

RESULTS

The mimivirus protein was identified as putative glycosyltransferase based on sequence similarities to prokaryotic and yeast glycosyltransferases. Based on BLAST analysis of the 251 amino acid protein L373 an N-terminal glycosyltransferase superfamily domain spanned residues 18 to 97 (Figure 1A). L373 was grouped within glycosyltransferase 32 family. Glycosyltransferase 32 family has more than 1'700 members of which 20 are characterized (9). Of these 20 glycosyltransferases the majority are eukaryotic and prokaryotic mannosyltransferases. In L373 an OCH1 multi-domain was predicted to be spanning residues 1 to 175 (Figure 1A). OCH1 is a yeast mannosyltransferase situated in the *cis*-Golgi apparatus. OCH1 initiates synthesis of poly-Man elongation of N-linked oligosaccharides of yeast glycoproteins (10). Other characterized members of glycosyltransferase 32 family are galactosyl-, N-acetylgalactosaminyl-, and two N-acetylglucosaminyltransferases. In L373 a capsular polysaccharide synthesis multi-domain was predicted to be spanning residues 5 to 136 (Figure 1A). This domain is found in a bi-specific α-GlcNAc/α-Glcl-transferase (Uniprot ID: Q4K231) encoded by *wcrL* which is involved in the synthesis of the capsular polysaccharide in *Streptococcus pneumonia* 70/86 (11, 12). The DXD motifs of L373 and WcrL align (Figure 1B). The DXD motif is involved in the coordination of a divalent metal ion in the catalytic pocket of many glycosyltransferases utilizing nucleotide-activated sugars as donor substrates (13).

The protein L373 was expressed as FLAG-fusion protein in *Sf-9* insect cells. L373 PCR product was cloned into pFastBac-FLAG(C) using BamHI and XbaI (Figure 2A). To control protein expression 5 µl post nuclear supernatant were used for an anti-

FLAG Western blot which showed a single band at 31.6 kDa corresponding to FLAG-L373 (Figure 2B). Sequence analysis suggested mannosyl- or N-acetyl-glucosaminyltransferase activity for L373. While GlcNAc is highly abundant in mimivirus only traces of Man are found (6). Post nuclear supernatant from *Sf-9* cells overexpression L373 or from *Sf-9* cells infected with the empty control vector was tested for nucleotide-sugar donor hydrolyzing activity. L373 showed a specific hydrolysis of UDP-GlcNAc with an average activity of 367.6 pmol/min/µg. L373 did not hydrolyze UDP-Glc or GDP-Man at a significant level in comparison to the negative control (Figure 3A). Glycosyltransferases with a DXD motifs are often dependent on divalent metal ions that are involved in stabilizing the catalytic pocket (13). In absence of $MnCl_2$ the UDP-GlcNAc hydrolysis activity of L373 was reduced to background levels (Figure 3B). We used a broad variety of acceptor substrates in combination with UDP-GlcNAc as nucleotide sugar donor to determine the glycosyltransferase activity of L373 (Table 1). We tested mimivirus total protein content and some known or predicted glycoproteins or peptides thereof ((14, 15), A. Hülsmeier, unpublished data). In addition we tested various pNP-monosaccharides that reflected the sugar content of mimivirus (5, 6). There was no glycosyltransferase activity detected on the tested proteins, peptides, and pNP-monosaccharides (data not shown).

DISCUSSION

We have shown that mimivirus L373 is a functional UDP-GlcNAc specific hydrolase with glycosyltransferase characteristics. L373 had homologies to both N-acetylglucosaminyl- and mannosyltransferases. The majority of glycosyltransferases, however, are donor-specific and do not use multiple nucleotide-sugar donors (16). This holds true for L373 as L373 did not hydrolyze GDP-Man. This finding goes in hand with the monosaccharide composition of mimivirus where GlcNAc is a major component

while only traces Man are found (5, 6). Sequence analysis of L373 identified a DXD motif within the predicted glycosyltransferase domain suggesting a dependence on divalent metal ions (13). Indeed, in absence of Mn^{2+} L373 hydrolysis activity was shed. We were able to rule out a variety of acceptor substrates of L373. Although a number of possible acceptors are known a broader screening of possible acceptor proteins and more complex carbohydrate acceptors would be necessary for L373 glycosyltransferase activity determination. Only a limited number of glycosylated proteins and peptides are known in mimivirus ((14, 15), A. Hülsmeier, unpublished data). It is likely that the sequon recognized by mimivirus glycosyltransferases differs from the canonical sequon. In *Paramecium bursaria* chlorella virus, another giant virus with host-independent glycosylation, the identified N-glycosylation sequon is unusual and differs from the eukaryotic sequon (17). It is therefore hard to predict possible mimivirus glycoproteins and glycosylation sites for core N-glycosyltransferases. From an O-glycan analysis in mimivirus we know that GlcNAc is mainly found in the terminal position of branched O-glycans (6). There is a limitation in availability of such complex glycans as suiting acceptor substrate for elongating glycosyltransferases. Most available pNP- or benzyl-coupled saccharides, which allow product purification, are monosaccharides or linear structures, but not complex, branched structures. L373 might, however, only be active as elongating enzyme if the core glycan is mature enough. It is common for glycosyltransferases to show strict acceptor specificity (18). As there is nothing known on mimivirus N-glycans and the location of the majority of mimiviral GlcNAc it is impossible to make further predictions on the glycosyltransferase activity of L373 to date.

DECLARATION OF INTEREST

The authors declare no conflict of interest with the content of this article.

FUNDING INFORMATION

This work was supported by the Research Credit of the University of Zurich to AR and by the Swiss National Foundation grant 310030_149949 to TH.

AUTHOR CONTRIBUTION STATEMENT

Anna Rommel and Thierry Hennet designed the experiments. Anna Rommel wrote the manuscript. Anna Rommel performed all experiments.

REFERENCES

1. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. Science. 2004;306(5700):1344-50. Epub 2004/10/16.

2. Kuznetsov YG, Xiao C, Sun S, Raoult D, Rossmann M, McPherson A. Atomic force microscopy investigation of the giant mimivirus. Virology. 2010;404(1):127-37. Epub 2010/06/17.

3. Piacente F, Bernardi C, Marin M, Blanc G, Abergel C, Tonetti MG. Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNA virus Mimivirus. Glycobiology. 2014;24(1):51-61. Epub 2013/10/11.

4. Parakkottil Chothi M, Duncan GA, Armirotti A, Abergel C, Gurnon JR, Van Etten JL, et al. Identification of an L-rhamnose synthetic pathway in two nucleocytoplasmic large DNA viruses. Journal of virology. 2010;84(17):8829-38. Epub 2010/06/12.

5. Piacente F, Marin M, Molinaro A, De Castro C, Seltzer V, Salis A, et al. The giant DNA virus Mimivirus encodes a pathway for the biosynthesis of the unusual sugar 4-amino-4,6-dideoxy-D-glucose (viosamine). The Journal of biological chemistry. 2011. Epub 2011/12/14.

6. Hulsmeier AJ, Hennet T. O-Linked glycosylation in Acanthamoeba polyphaga mimivirus. Glycobiology. 2014;24(8):703-14. Epub 2014/05/06.

7. Luther KB, Hulsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennet T. Mimivirus collagen is modified by a bifunctional lysyl hydroxylase and glycosyltransferase enzyme. The Journal of biological chemistry. 2011. Epub 2011/11/03.

8. Hennet T, Dinter A, Kuhnert P, Mattu TS, Rudd PM, Berger EG. Genomic cloning and expression of three murine UDP-galactose: beta-N-acetylglucosamine beta1,3-galactosyltransferase genes. The Journal of biological chemistry. 1998;273(1):58-65. Epub 1998/02/07.

9. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic acids research. 2014;42(Database issue):D490-5. Epub 2013/11/26.

10. Nakayama K, Nagasu T, Shimma Y, Kuromitsu J, Jigami Y. OCH1 encodes a novel membrane bound mannosyltransferase: outer chain elongation of asparagine-linked oligosaccharides. The EMBO journal. 1992;11(7):2511-9. Epub 1992/07/01.

11. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS genetics. 2006;2(3):e31. Epub 2006/03/15.

12. Oliver MB, Jones C, Larson TR, Calix JJ, Zartler ER, Yother J, et al. Streptococcus pneumoniae serotype 11D has a bispecific glycosyltransferase and expresses two different capsular polysaccharide repeating units. The Journal of biological chemistry. 2013;288(30):21945-54. Epub 2013/06/06.

13. Ramakrishnan B, Boeggeman E, Ramasamy V, Qasba PK. Structure and catalytic cycle of beta-1,4-galactosyltransferase. Current opinion in structural biology. 2004;14(5):593-600. Epub 2004/10/07.

14. Klose T, Herbst DA, Zhu H, Max JP, Kenttamaa HI, Rossmann MG. A Mimivirus Enzyme that Participates in Viral Entry. Structure. 2015;23(6):1058-65. Epub 2015/05/20.

15. Boyer M, Azza S, Barrassi L, Klose T, Campocasso A, Pagnier I, et al. Mimivirus shows dramatic genome reduction after intraamoebal culture. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(25):10296-301. Epub 2011/06/08.

16. Rini J, Esko J, Varki A. Glycosyltransferases and Glycan-processing Enzymes. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

17. De Castro C, Molinaro A, Piacente F, Gurnon JR, Sturiale L, Palmigiano A, et al. Structure of N-linked oligosaccharides attached to chlorovirus PBCV-1 major capsid protein reveals unusual class of complex N-glycans. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(34):13956-60. Epub 2013/08/07.

18. Patenaude SI, Seto NO, Borisova SN, Szpacenko A, Marcus SL, Palcic MM, et al. The structural basis for specificity in human ABO(H) blood group biosynthesis. Nature structural biology. 2002;9(9):685-90. Epub 2002/08/29.

FIGURE LEGENDS

FIGURE 1. (A) BLAST analysis of the 251 amino acid protein L373 with predicted domains. The glycosyltransferase (GT) sugar binding region was predicted to be spanning residues 18 to 97, the OCH1 domain was predicted to be spanning residues 1 to 175, and the capsular polysaccharide synthesis protein domain was predicted to be spanning residues 5 to 136. (B) Multiple sequence alignment of OCH1, L373, and WcrL. Identical residues are shaded in black. Similar residues are shaded in gray. DXD motif is marked by dashed box.

FIGURE 2. (A) Vector map of pFastBac-FLAG(C)-L373. XbaI and BamHI were used for cloning of L373 PCR product. The plasmid contained a gentamycin resistance (GmR), an ampicillin resistance (AmpR) and the bacterial transposons Tn7R/ Tn7L. (B) Anti-FLAG Western blot of *Sf-9* post nuclear cell lysate from *Sf-9* cells overexpressing L373. The band at 31.6 kDa corresponds to FLAG-L373.

FIGURE 3. (A) Hydrolysis activity of L373 using UDP-Glc, UDP-GlcNAc, and GDP-Man *in vitro*. Grey bars represent *Sf-9* cells infected with empty control vector (pFB_empty). Open bars represent *Sf-9* cells overexpressing L373 (pFB_L373). (B) UDP-GlcNAc hydrolysis activity of L373 in presence and absence of Mn^{2+}. Two independent experiments with triplicates in each condition were done. Results are shown as mean ± SEM.

TABLE 1. Acceptor substrates used to identify glycosyltransferase activity of L373.

Type of acceptor substrate	Mimivirus ORF number or chemical name
Full length proteins	L425, R459, R856
Peptides	L357, L410, R135, R459
pNP-monosaccharides	pNPαFuc, pNPαGal, pNPα/βGlc, pNPα/βGlcNAc, pNPαRha, pNPαXyl

Figure 1

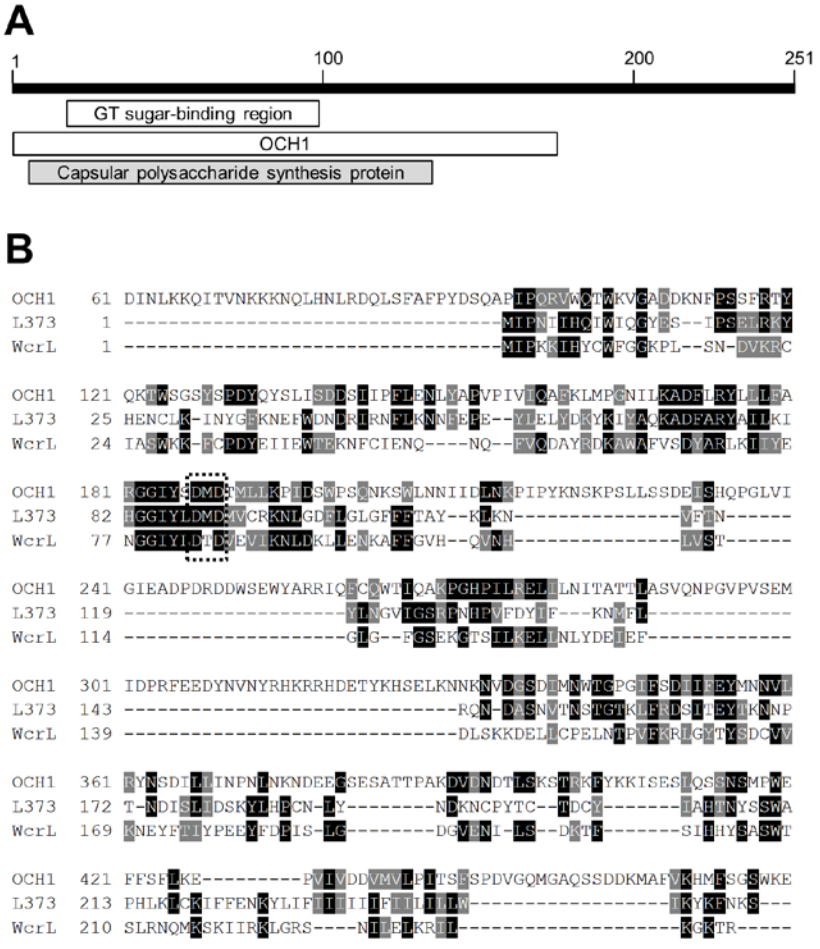

A
1
100
200
251
GT sugar-binding region
OCH1
Capsular polysaccharide synthesis protein
B
OCH1 61 DINLKKQITVNKKKNQLHNLRDQLSFAFPYDSQAPIPQRVWQTWKVGALDKNFPSSFRTY
L373 1 ----------MIPNTIHQIWIQGYES--IPSELRKY
WcrL 1 ----------MIPKKIHYCWFGGKPL--SN-DVKRC
OCH1 121 QKTWSGSYSPDYQYSLISDDSLIIPFLENLYAPVPIVIQAEKLMPGNILKADFLRYLLLEA
L373 25 HENCLK-INYGEKNEFWDNDRIRNFLKNNEEPE--YLELYDKYKIYAQKADFARYAIIKI
WcrL 24 IASWKK-FGPDYEIIEWTFKNFCIENQ----NQ--FVQDAYEDKAWAFVSDVARLKIIYE
OCH1 181 RGGIYSDMDSMLIKPLDSWPSQNKSWLNNIIDLNKPIPYKNSKPSLLSSDEISHQPGLVI
L373 82 HGGIYLDMDIVCRKNLGDLLGLGFFTAY--KLKN----------VFIN------
WcrL 77 NGGIYIDTDIEVIKNLDKLLENKAFFGVH--QVNH----------LVST------
OCH1 241 GIEADPDRDDWSEWYARRIQECQWTIQAKPGHPILKELLLNITATTLASVQNPGVPVSEM
L373 119 ----------YLNGVIGSFPNHPVFDYIF---KNMFL----------
WcrL 114 ----------GLG--FGSEKGTSILKELLNLYDEIEF----------
OCH1 301 IDPRFEEDYNVNYRHKRRHDETYKHSELKNNKNVDGSDIMNWTGPGIESDIIFEYMNNVL
L373 143 ----------RQN-DASNVTNSTGTKLFRDSILEYIKNNP
WcrL 139 ----------DLSKKDELICPELNTPVFKRIGYTYSDGVV
OCH1 361 RYNSDILLINPNLNKNDEPGSESATTPAKDVDNDTLSKSTRKEYKKISESIQSSNSMFWE
L373 172 T-NDISLIDSKYLHECN-EY--------NDKNCPYTC--TDCY-------IAHINYSSWA
WcrL 169 KNEYFTIYPEEYFDEIS-LG--------DGVENI-LS--DKTG-------SIIHYSASWT
OCH1 421 FFSELKE---------PVLVDDVMVLPITSESPDVGQMGAQSSDDKMAFVKHMESGSWKE
L373 213 PHLKLCKIFFENKYLIFTIIIIFLILILW----------IKYKFNKS---
WcrL 210 SLRNQMKSKIIRKLGRS----NIIELKRIL----------KGKTR-----

Figure 2

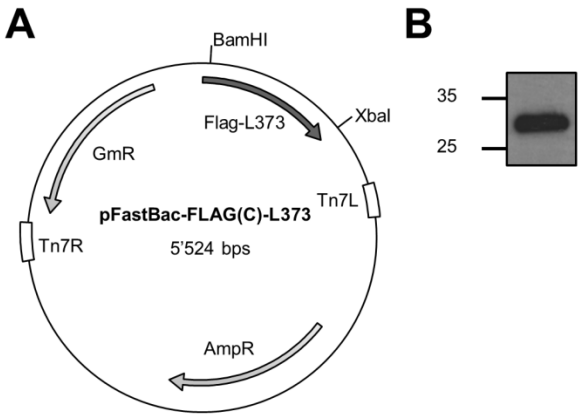
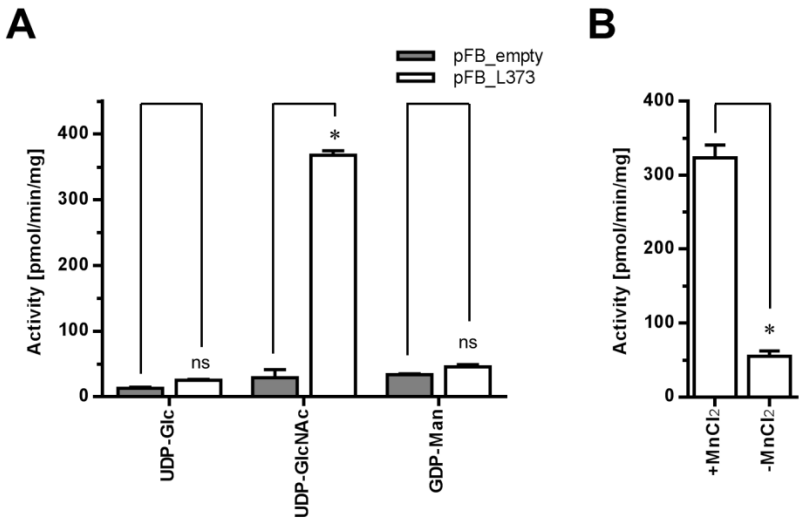

A
BamHI
Flag-L373
XbaI
GmR
Tn7L
pFastBac-FLAG(C)-L373
Tn7R
5'524 bps
AmpR
B
35
25

Figure 3


A
pFB_empty
pFB_L373
Activity [pmol/min/mg]
400
300
200
100
0
ns
*
ns
UDP-Glc
UDP-GlcNAc
GDP-Man
B
Activity [pmol/min/mg]
400
300
200
100
0
*
+MnCl2
-MnCl2

ADDITIONAL RESULTS

CLONING, EXPRESSION, AND ACTIVITY SCREENING OF PUTATIVE MIMIVIRUS GLYCOSYLTRANSFERASES

EXPERIMENTAL PROCEDURES

Expression and purification of recombinant glycosyltransferase candidates

Recombinant L193, R139, R363, and R654 were expressed in *E. coli* BL21 Rosetta as His_{10}-fusion protein in pET16b or pET28a (Merck, Darmstadt, Germany). Viral DNA was purified using phenol/ chloroform (Roth, Karlsruhe, Germany). The DNA sequences corresponding to mimivirus L193, R139, R363, and R654 (GenBank Accession number: NC_014649.1) were amplified using the primers listed in table 5. The amplification primers incorporated a restriction endonuclease site (New England Bioscience, Hitchin, UK) (Table 5). For protein expression 100 ml lysogeny broth containing 100 μg/ml ampicillin or 50 μg/ml kanamycin (both from Sigma-Aldrich, Buchs, Switzerland) were inoculated with 1 ml from an overnight culture. Protein expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (Biosolve, Dieuze, France) at OD_{600} of 0.4-0.6. After induction *E. coli* were grown shaking as follows: L193 at 16°C for 22 h, R139 at 30°C for 1.5 h, R363 at 16°C for 6 h, and R654 at 30°C for 1.5 h, then pelleted at 6'000 x g, resuspended in 15 ml 500 mM NaCl, 20 mM Tris pH 7.4, 10% v/v glycine, 10 mM imidazole (MCAC10) and lysed by an Emulsiflex C5 French press (Avestin, Mannheim, Germany). His-tagged proteins were purified from the soluble fraction using 100 μl Ni-Sepharose 6 Fast Flow beads (GE Healthcare, Glattbrugg, Switzerland). After binding to beads in MCAC10 at 4°C for 4 h, rotating, proteins were eluted with 400 mM imidazole (MCAC400).

Table 5 Primers, vectors, and restriction enzymes used for cloning of putative mimivirus glycosyltransferases

ORF	Forward primer (5′-3′)	Reverse primer (5′-3′)	PCR product (bp)	Vector (restriction enzymes)
R139	ATAGAATTCATGTCCAATCAAGCAACATCTG	AATACTCGAGATTCTTTATGGAACCCTCCTCA	763	pET28 (EcoRI, XhoI)
L193	CGCAAGCTTGATTCCTGTTACAATCCTCAC	CGCCTCGAGTGGAGTTTCAAACAGATTAATC	1′841	pET28 (HindIII, XhoI)
R363	ATACAAGCTTATGCAAGAAATAAATGGCACATCTC	AATCTCGAGTTCAATACATTCCACGTACTCTTG	787	pET28 (HindIII, XhoI)
R654	ATACTCGAGGAAAGCTGTAAAATTATTTGTA	ATAGGATCCTTAAATGTATAATGTTTGAATT	1′871	pET16 (XhoI, BamHI)

Control of Protein Expression

For protein separation 15 µl of purified enzyme were subjected to 10% SDS-PAGE. Proteins were transferred to nitrocellulose membrane (GE Healthcare) in semi-dry manner. After transfer the membrane was blocked in 5% milk in PBS-Tween (0.01%) (PBS-T) at room temperature for 1 h. Primary anti-His antibody (Sigma-Aldrich) was diluted 1:5′000 in PBS-T containing 1% bovine serum albumin (BSA) (Sigma-Aldrich) and either incubated at room temperature for 1 h or at 4°C overnight. After antibody incubation the membrane was washed three times for 5 min in PBS-T followed by incubation with secondary antibody, peroxidase conjugated goat anti-mouse IgG (Sigma-Aldrich, Buchs, Switzerland) diluted 1:10′000 in PBS-T/ 1% BSA at room temperature for 1 h. After antibody incubation the membrane was washed three times

for 5 min in PBS-T followed by development with SuperSignal chemiluminescent substrate (Thermo Scientific, Waltham, MA).

Determination of activity

Glycosyltransferase activity was measured in 20 mM Tris-HCl pH 7.4 containing 0.5 µl 20 µCi/ml UDP-[^{14}C]-Glc or UDP-[^{14}C]-GlcNAc or GDP-[^{3}H]-Fuc (Perkin-Elmer, Waltham, MA or American Radio Chemicals, St. Louis, MO), 20-50 µM UDP-sugar (Sigma-Aldrich or Carbosynth, Compton, UK), 10 mM $MnCl_2$ (Sigma-Aldrich), 1 mM DTT (Sigma-Aldrich), 10-50 mM 4-nitrophenyl (pNP)-sugar in DMSO (Sigma-Aldrich) as acceptor substrate or 20 µg mimivirus peptides (GenScript, Piscataway Township, NJ), and 10 µg of purified protein. The UDP-sugars tested as donor substrates were UDP-Glc, UDP-GlcNAc, GDP-Fuc, UDP-Rha, and UDP-Vio. The pNP-sugars tested as acceptor substrates were pNP-α/β-Glc, pNP-α/β-GlcNAc, and pNP-α-Rha (Sigma-Aldrich). Assays were incubated at 35°C for 4 h. Assay containing pNP-sugars were stopped by addition of ice cold 500 µl H_2O, and purified over SepPak-C18 cartridges (Waters, Milford, MA). Samples were added to the cartridges, washed with 10 ml of H_2O, and eluted with 5 ml of methanol (Sigma-Aldrich). Radioactivity was measured after addition of 10 ml of IRGASAFE scintillation liquid (Perkin-Elmer) in a Tri-Carb 2900TR scintillation counter (Perkin-Elmer). Assays containing mimivirus peptides were stopped by addition of 500 µl 5% TCA / 5% phosphotungstic acid and incubated on ice for 30 min. Precipitates were recovered on 1 µm glass filters (VWR, Radnor, PA) using a vacuum manifold and radioactivity was measured as described above. Assays containing UDP-Vio or UDP-Rha (kindly provided by M. Tonetti) were incubated under the same conditions without the supplementation of radioactively labeled nucleotide sugar donor. The release of UDP was measured using the UDP-Glo kit (Promega, Madison, WI) following the manufacturer's manual.

RESULTS

Mimivirus L193 is a 601 amino acid protein with a molecular weight of 70.8 kDa. BLAST analysis predicted the presence of two glycosyltransferase domains: an N-terminal GT-A domain and a C-terminal GT-B domain (Figure 12). The GT-A domain of L193 showed sequence homologies to GT 2 family proteins. GT 2 has almost 70'000 family members of which 269 are characterized (1, 2). These glycosyltransferases are all inverting enzymes, but show a huge variety in terms of functionality and are found amongst archaea, bacteria, and eukaryotes. In L193 a WcaA multi-domain was predicted to be spanning residues 1 to 280 including the GT-A domain. WcaA is involved in the synthesis of colanic acid in *E. coli*. Colanic acid is a polyanionic heteropolysaccharide containing a repeat unit with D-Glc, L-Fuc, D-Gal, and D-glucuronate sugars that are nonstoichiometrically decorated with O-acetyl and pyruvate side chains (3). The GT-B domain showed sequence homologies to GT 1 family proteins. GT 1 contains more than 9'700 members of which 366 are characterized (1, 2). These glycosyltransferases show a huge variety in terms of functionality and are found amongst bacteria and eukaryotes. An RfaG multi-domain was predicted to be spanning the GT-B domain. RfaG is involved in the core synthesis of *E. coli* LPS (4). Phyre[2] predicted a GT 1 family domain in the C-terminus of L193 with a UDP-glycosyltransferase/glycogen phosphorylase fold and homologies to starch synthase.

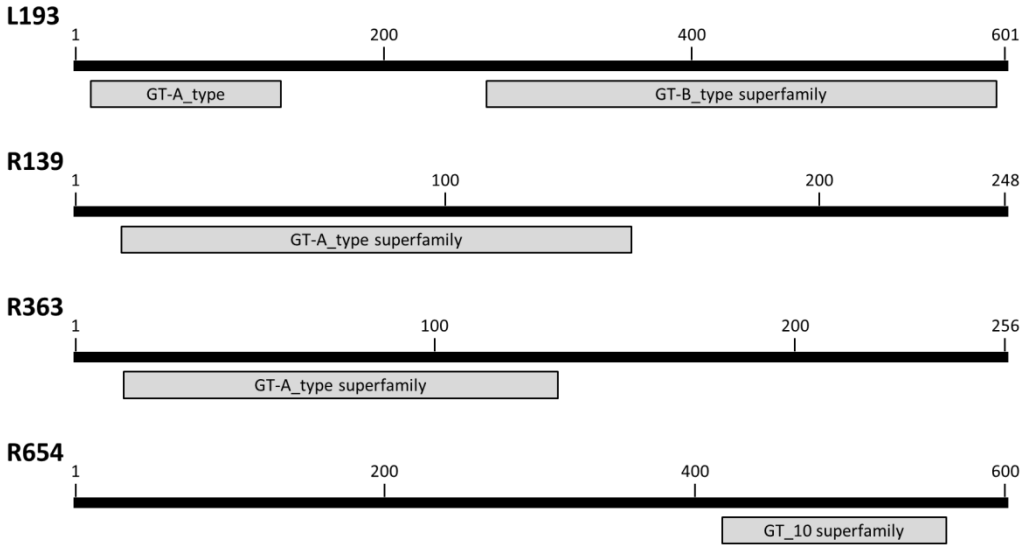


Figure 12 BLAST analysis of mimivirus ORFs L193, R139, R363, and R654with predicted domains. L193 was predicted to have an N-terminal glycosyltransferase (GT)-A type domain and a C-terminal GT-B type domain. R139 and R363 were predicted to have an N-terminal GT-A type domain. R654 was predicted to have a C-terminal GT-10 domain.

Mimivirus R139 is a 248 amino acid protein with a molecular weight of 28.9 kDa. BLAST analysis predicted an N-terminal GT-A domain (Figure 12). R139 showed sequence homologies to prokaryotic glycosyltransferases. BLAST search, however, resulted only in uncharacterized GT 2 family proteins or hypothetical proteins. Mimivirus R139 was found in the same locus as two nucleotide-diphospho sugar synthetizing enzymes of mimivirus, namely L136, an enzyme involved in the synthesis of UDP-Vio, and L141, an enzyme involved in the synthesis of UDP-Rha (5, 6). Phyre^2 analysis of R139 predicted a nucleotide-diphospho sugar transferase fold, possibly being a polypeptide N-acetyl-galactosaminyltransferase, a chondroitin polymerase, or a mannosyl-3-phospho-glycerate synthase.

Mimivirus R363 is a 256 aa protein with a molecular weight of 29.4 kDa. BLAST analysis predicted an N-terminal GT-A domain (Figure 12). R363 showed sequence homologies

with giant virus, cyanobacterial, and archaeal glycosyltransferases. BLAST search, however, resulted only in uncharacterized glycosyltransferases or hypothetical proteins. R363 was grouped within GT 2 family. Phyre[2] analysis predicted R363 to be a chondroitin synthase with an N-acetyl-galactosaminyltransferase activity.

Mimivirus R654 is a 600 aa protein with a molecular weight of 70.7 kDa. BLAST analysis predicted C-terminal GT 10 superfamily domain (Figure 12). R654 showed sequence homologies with both prokaryotic and eukaryotic fucosyltransferases throughout the glycosyltransferase domain. R654 was grouped within GT 10 family. Glycosyltransferase 10 has more than 500 members of which 45 are characterized (1, 2). All members of GT 10 contain a GT-B fold, work in an inverting way, and have catalytic activity as prokaryotic or eukaryotic fucosyltransferases forming either α1,3- or α1,4-linkages. Phyre[2] analysis predicted domain and secondary structure homologies to fucosyltransferase -like glycosyltransferases.

L193, R139, and R363 were expressed in *E. coli* BL21 Rosetta as C-terminal His_{10}-fusion proteins; R645 was expressed in *E. coli* BL21 Rosetta as N-terminal His_{10}-fusion protein. To control protein expression and purification 15 µl proteins in MCAC400 were subjected to SDS-page followed by an anti-His Western blot. All glycosyltransferase candidates were detected at the expected protein size (Figure 13).

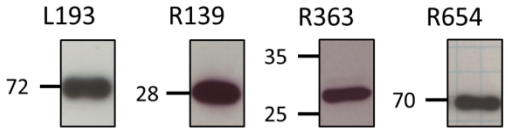


Figure 13 Anti-His Western blot of recombinantly expressed and purified L193 (70.8 kDa), R139 (28.9 kDa), R363 (29.4 kDa), and R654 (70.7 kDa). Single bands correspond to the predicted protein sizes.

Sequence analysis of L193 did not give clear hints towards the functionality of L193. Therefore we did a broad screening with the main monosaccharides found in mimivirus as nucleotide sugar donor and a variety of both peptide and monosaccharide acceptors (5-8). We did not detect any L193 glycosyltransferase activity within the conditions tested (Table 6).

Table 6 Tested combinations of donor nucleotide sugars and acceptor substrates and L193 as putative glycosyltransferase *in vitro* X indicates the absence of glycosyltransferase activity.

	Donor nucleotide sugar			
Acceptor substrate	**UDP-Glc**	**UDP-GlcNAc**	**UDP-Vio**	**UDP-Rha**
L357	X	X	X	X
L410	X	X	X	X
R135	X	X	X	X
R459	X	X	X	X
pNP-Glc (α or β) or Glc	X	X	X	X
pNP-GlcNAc (α or β) or GlcNAc	X	X	X	X
pNPαRha or Rha	X	X	X	X

Sequence analysis of R139 did not give any clear hints towards the functionality of R139. As member of the same gene cluster like two enzymes involved in the synthesis of UDP-Rha (R141) and UDP-Vio (L136), we focused on using UDP-Vio and UDP-Rha as donor nucleotide sugars in a first round of activity tests (5, 6). Since we did not detect any glycosyltransferase activity using UDP-Rha and UDP-Vio, we tested UDP-Glc and UDP-GlcNAc as nucleotide sugar donors. We did not detect any R139 glycosyltransferase activity within the conditions tested (Table 7).

Table 7 Tested combinations of donor nucleotide sugars and acceptor substrates and R139 as putative glycosyltransferase *in vitro*. X indicates the absence of glycosyltransferase activity.

	Donor nucleotide sugar			
Acceptor substrate	**UDP-Glc**	**UDP-GlcNAc**	**UDP-Vio**	**UDP-Rha**
L357	X	X	X	X
L410	X	X	X	X
R135	X	X	X	X
R459	X	X	X	X
pNPαGlc	X	X	X	X
pNPβGlc	X	X	X	X
pNPαGlcNAc	X	X	X	X
pNPαRha	X	X	X	X

Sequence analysis of R363 did not give any clear hints towards the functionality of R363. Therefore we did a broad screening with the main monosaccharides found in mimivirus as donor nucleotide sugar and a variety of both peptide and monosaccharide acceptors. We did not detect any R363 glycosyltransferase activity within the conditions tested (Table 8).

Table 8 Tested combinations of donor nucleotide sugars and acceptor substrates and R363 as putative glycosyltransferase *in vitro*. X indicates the absence of glycosyltransferase activity.

	Donor nucleotide sugar			
Acceptor substrate	**UDP-Glc**	**UDP-GlcNAc**	**UDP-Vio**	**UDP-Rha**
L357	X	X	X	X
L410	X	X	X	X
R135	X	X	X	X
R459	X	X	X	X
pNP-Glc (α or β) or	X	X	X	X

Glc				
pNP-GlcNAc (α or β) or GlcNAc	X	X	X	X
pNPαRha or Rha	X	X	X	X

Sequence analysis suggested fucosyltransferase activity for R654. Therefore, we used GDP-Fuc and various acceptors to test for R654 activity. As we did not identify any specific activity we continued with a broad screen with the main monosaccharides found in mimivirus as donor nucleotide sugar and a variety of both peptide and monosaccharide acceptors We could not detect any R654 glycosyltransferase activity in the conditions tested (Table 9).

Table 9 Tested combinations of donor nucleotide sugars and acceptor substrates and R654 as putative glycosyltransferase *in vitro*. X indicates the absence of glycosyltransferase activity.

	Donor nucleotide sugar				
Acceptor substrate	**GDP-Fuc**	**UDP-Glc**	**UDP-GlcNAc**	**UDP-Vio**	**UDP-Rha**
L357	X	X	X	X	X
L410	X	X	X	X	X
R135	X	X	X	X	X
R459	X	X	X	X	X
pNP-Glc (α or β) or Glc	X	X	X	X	X
pNP-GlcNAc (α or β) or GlcNAc	X	X	X	X	X
pNPαRha or Rha	X	X	X	X	X

DISCUSSION

We recombinantly expressed and purified the four putative mimivirus glycosyltransferases L193, R139, R363, and R654. We could rule out a number of possible glycosyltransferase activities for all candidates. The viral proteins were recombinantly expressed in *E. coli* BL21 Rosetta but not in any other *E. coli* strain. The BL21 Rosetta strain is designed to enhance expression of eukaryotic proteins. They supply codons rarely used by *E. coli* thus providing for universal translation (9-12). It is likely that the codons naturally provided by *E. coli* are not sufficient for translation of mimivirus proteins because many of the mimivirus proteins have eukaryotic origin or a common evolutionary background with eukaryotic proteins (13).

L193, R139, R363, and R654 were not active under the tested conditions *in vitro*. This could be due to various reasons. Many glycosyltransferases have both N- and O-glycosylation sites themselves (14). The consensus glycosylation sequences of eukaryotic and prokaryotic glycosylation sites are not necessarily the same as in giant viruses. Due to recombinant expression in *E. coli*, the mimivirus glycosyltransferases would lack this glycosylation necessary for their activity.

Most glycosyltransferases work in a strict acceptor specific manner. For example, the human B blood group α1,3 galactosyltransferase is only active if the Gal residue on the acceptor substrate is linked to Fuc in an α1,2 linkage and not if the Gal residue is modified by other monosaccharides (15). The commercial availability of complex glycans as possible acceptors substrates is limited, even with a number of O-glycan structures known (8).

In mimivirus, various glycans are further modified by methylation. Methylation is described for GlcNAc, Rha, and Vio (8). Methylation is mostly performed as the final modification step after biosynthesis of the glycan. Mycobacteria, however, express a rhamnosyltransferase that transfers methylated Rha to the oligosaccharide acceptor

chain (16). If mimivirus glycosyltransferases have a preference towards the methylated donor nucleotide sugar detection of glycosyltransferase activity *in vitro* using non-methylated donor nucleotide sugar is impossible.

Little is known about the composition of the virus factory and we do not know whether the presence of certain co-factors or complex partners is necessary for proper functioning of mimivirus glycosyltransferases. Cellulose biosynthesis in bacteria for example is catalyzed by a protein complex consisting of a glycosyltransferase (BcsA), a membrane-associated periplasmic protein (BcsB), and an outer membrane protein (BcsC) (17).

REFERENCES

1. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic acids research. 2014;42(Database issue):D490-5. Epub 2013/11/26.

2. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnology for biofuels. 2013;6(1):41. Epub 2013/03/22.

3. Garegg PJ, Lindberg B, Onn T, Sutherland IW. Comparative structural studies on the M-antigen from Salmonella typhimurium. Escherichia coli and Aerobacter cloacae. Acta chemica Scandinavica. 1971;25(6):2103-8. Epub 1971/01/01.

4. Parker CT, Pradel E, Schnaitman CA. Identification and sequences of the lipopolysaccharide core biosynthetic genes rfaQ, rfaP, and rfaG of Escherichia coli K-12. Journal of bacteriology. 1992;174(3):930-4. Epub 1992/02/01.

5. Piacente F, Marin M, Molinaro A, De Castro C, Seltzer V, Salis A, et al. The giant DNA virus Mimivirus encodes a pathway for the biosynthesis of the unusual sugar 4-amino-4,6-dideoxy-D-glucose (viosamine). The Journal of biological chemistry. 2011. Epub 2011/12/14.

6. Parakkottil Chothi M, Duncan GA, Armirotti A, Abergel C, Gurnon JR, Van Etten JL, et al. Identification of an L-rhamnose synthetic pathway in two nucleocytoplasmic large DNA viruses. Journal of virology. 2010;84(17):8829-38. Epub 2010/06/12.

7. Piacente F, Bernardi C, Marin M, Blanc G, Abergel C, Tonetti MG. Characterization of a UDP-N-acetylglucosamine biosynthetic pathway encoded by the giant DNA virus Mimivirus. Glycobiology. 2014;24(1):51-61. Epub 2013/10/11.

8. Hulsmeier AJ, Hennet T. O-Linked glycosylation in Acanthamoeba polyphaga mimivirus. Glycobiology. 2014;24(8):703-14. Epub 2014/05/06.

9. Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. Current opinion in biotechnology. 1995;6(5):494-500. Epub 1995/10/01.

10. Kurland C, Gallant J. Errors of heterologous protein expression. Current opinion in biotechnology. 1996;7(5):489-93. Epub 1996/10/01.

11. Brinkmann U, Mattes RE, Buckel P. High-level expression of recombinant genes in Escherichia coli is dependent on the availability of the dnaY gene product. Gene. 1989;85(1):109-14. Epub 1989/12/21.

12. Baca AM, Hol WG. Overcoming codon bias: a method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. International journal for parasitology. 2000;30(2):113-8. Epub 2000/03/08.

13. Yutin N, Wolf YI, Koonin EV. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology. 2014. Epub 2014/07/22.

14. Varki A, Esko JD, Colley KJ. Cellular Organization of Glycosylation. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

15. Rini J, Esko J, Varki A. Glycosyltransferases and Glycan-processing Enzymes. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

16. Staudacher E. Methylation--an uncommon modification of glycans. Biological chemistry. 2012;393(8):675-85. Epub 2012/09/05.

17. Gloster TM. Advances in understanding glycosyltransferases from a structural perspective. Current opinion in structural biology. 2014;28:131-41. Epub 2014/09/23.

GENOME ENGINEERING IN MIMIVIRUS USING THE CRISPR-CAS9 SYSTEM IN *ACANTHAMOEBA*

EXPERIMENTAL PROCEDURES AND RESULTS

To develop a tool for the knock-out of single or multiple mimivirus genes we adapted the clustered regularly interspaced short palindromic repeats (CRISPR) adaptive immune system from microbes which used for genome engineering in a broad range of species (1-3). In order to produce knock-outs of mimivirus genes we manipulated a CRISPR plasmid for *Acanthamoeba* specific expression of Cas9 and gRNA. Transfection with the *Acanthamoeba* specific CRISPR plasmid would allow us to target any mimivirus gene by cloning of a mimivirus specific gRNA into the plasmid. We used the pSpCas9 (BB)-2A-Puro (PX459) plasmid (Addgene plasmid # 62988, gifted from Feng Zhang) which was designed for expression in mammalian cells and has a FLAG-tagged Cas9 from *Streptococcus pyogenes* with a C-terminal puromycin resistance and a cloning site for gRNA (Figure 14A) (3). We modified this in plasmid in three steps to generate a plasmid for transfection, Cas9 and gRNA expression in *Acanthamoeba* (Figure 14).

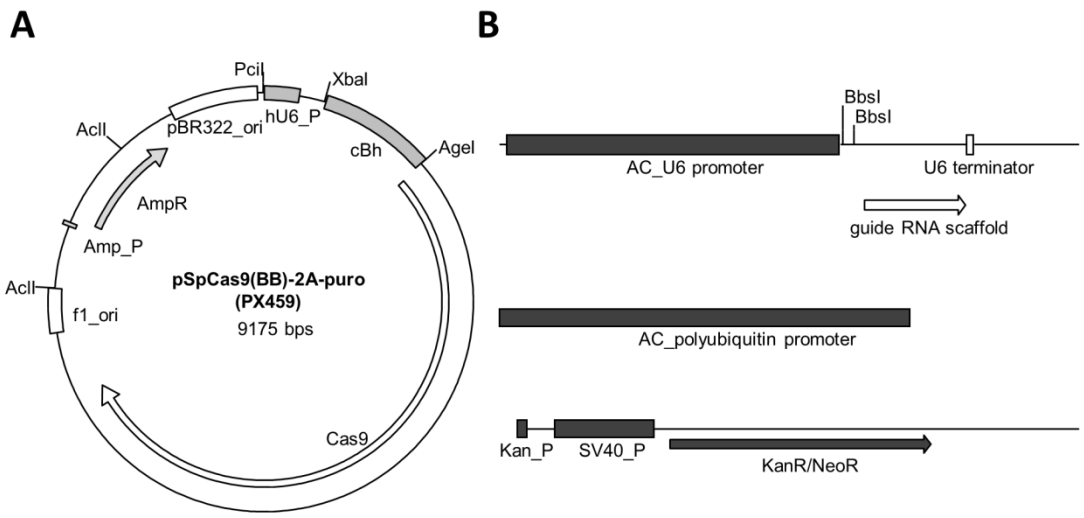


Figure 14 A Vector map of pSpCas9 (BB)-2A-Puro (PX459). The exchanged cassettes (mammalian promoter cassettes and ampicillin cassette) are shaded in light gray. The plasmid contains an ampicillin resistance cassette (AmpR) with the necessary ampicillin promoter (Amp_P). PciI, XbaI, AgeI, AclI mark restriction enzyme sites that were used for cassette exchange. **B. Newly introduced cassettes tailored for gRNA, Cas9 expression, and selection in *Acanthamoeba*.** *Acanthamoeba castellanii* U6 spliceosomal RNA promoter (AC_U6 promoter) and *Acanthamoeba castellanii* polyubiquitin promoter (AC_polyubiquitin promoter) were cloned into PX459 to drive amoeba specific gRNA and Cas9 expression. BbsI restriction enzyme sites were used for introduction of gRNA. For selection in bacteria and amoeba a kanamycin/neomycin resistance cassette (KanR/NeoR) containing the bacterial (Kan_P) and amoebal promoter (SV40_P) was introduced.

In a first step we exchanged the human U6 promoter driving gRNA expression with the corresponding sequence from *Acanthamoeba*. In 2008 a U6 spliceosomal RNA sequence was identified in *Acanthamoeba castellanii* (4). Using this sequence we analyzed the *Acanthamoeba* genome to target the upstream promoter region (Figure 15A) (5). We had the predicted promoter plus the gRNA insertion site, the gRNA scaffold and the U6 terminator synthesized to facilitate cloning (GenScript, Piscataway Township, NJ)

(Figure 14B). The whole cassette was introduced into the PX459 plasmid using PciI and XbaI (New England Bioscience, Hitchin, UK).

A

GGTCCGTTGATGCAGTAGTTCATCACGTGTAGGCCAAAAAGAAAATCTCTCCAAATTCACGCCAAACAAACTTGTG
TCCCCAAACTTTCACTGCCAAACTAGCCGTTAATGTCGACGGATAATTAAGTGAGGTGGGCGGTAACTTCCCGGAT
TGATAGATGCAATCCGGGATTATGTCCGGGTTGCAAAACGTGGTAACGGTCACGGAGCGGTGGCGTTCGCGCGTTG
TGGCGAGGACGAGAGAATGGTCGCGCACCTGCGCTTTTGTTCTGGCGTGTGGGTGGAGGCAACCCAACGGTCACCC
AGCCGTTGCTGAAATTGGCTAATTCACTTTCGGACTTGGACTTGTGGTAGCGGTCTTGCCCGCCGTCCACGTTCAC
CACCTCACCACCACAAACTGCTTAGACCCGGTTTTGCTGATACAGAGCCAACACCGCGCCGCAAGGCACAAGGATT
GTTCTTCTATTCTTTGGTTTCTTTATTGCCCTTTTTTTCCATCCCACAAGCGGAGAGGGGGTGGCGGGAGAAGAGG
TTGAAGAAGGCCGAGGCGGCGACGTACAAGGTCGGACCTCCCGAAAGGGAGGCTCCATCTGTTAAAATTGGAACGA
TACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCAAAATCGAGAAGATACCCAACTTTTTT

B

CGAGCTCTAGCACAAACTCGGCCACGTCGACCCCGTCGGGTCCGGACGTCGAGCGAAAGCTGAGCGAGTGCTCGAT
CGGCGCCAGGCTCAGGCTATAGTCGGCGTCGGCCAGGAGCACGACGTCGAACACCTCCACCTTGGCCACTGCGCGA
CGAGGCTCGCGTGGACGACGACGACGAGCGAATGGGCAGCGGGACGTAGTAGGTGGTCATGTTGGCCTCGGGTACG
CCGCGCTGGTGGGTGATGCTGGCGTTGAGGTCGAGCGGGGCCGCCAAGGAAGAATTAGCGGCCAAGACACTCGTGA
CACGATATCGAACCGTTGTGTGCTCGACGGCCGACGATGAGGAATTGGTAGATGAAAGGCGGAGCGACCACTTGAA
GCTATGGGGCGCGATGCCGAACGAGCGACCGAAGAAGTTGCCTCGCTCCGGCCTCCGGTGGAACGAGATCTCCTGT
TCGATCGCCGTCGCGTCGGCGTTGGCCACATAGGCAAAGGTCGTCCGAGCGTCGGAGTCTCCGCTGGACTCGTTCA
GCTGGGTTCGCTGGAAGCCGGTCAACGATGAGGCCGGAACCGTGATATCGCCCACTTGGATGCCTTCCAGCTCGGA
GGCAGTGTCAGCACTGCCTCCTCTTCGCTGCCTGACGTCGTGTCATTCCCTACAACACACACAAACACACAAGATC
AGGTGAGCCGACGAATCACGCTGACGATGAGCGAGGCCTACCGCGAATGGTGGCGGTGAGGTTGAAGCTGGCGGCG
TCGGCTCCGAAGCTCGACGTCACGTTGACCACCTGCAGCCCGGCAGCTTGATGCCGGGCAATGGCGACGGCGACGG
CGACGGCGACGGCATTGAGGAGTTGGACTGGTCCGGCGCGCAGCCGCTGACCGTCAGCTTGAAGCTGTTGCCGCCG
TCAACCACCCATGACTGGTTCGGGTTGGCCCGGTAGGCCAGGCCTAGCAGCACCACCCGTGCAGAGCCCGCGGGAC
TCGAGTCCACGGCCAGGCTCCCAGTGCGCGTGGACCCGGCATCGACCACGAGCCAATAGAGGCCAGGCCCAAGCAC
TAAGTCATCGGCAGTTGCGGAGGTCGGCGTTGCCGTGAAGCCGACGATTTCTGGTGTGACGTCCAGCGTGGCGTTG
GGCACGGCGGTGGCTGGGTGCGCCTGCGTTATCGGCGTAGAGCCGGCCGAATCCACGCCTCTGGGCCCTGTCCACA
GGCGCACAGTCGCGATCAGCCCACAAAAGAACGAAACGAATAAAGAACGAAATAAAAGAAAGCGAAGAGAAGCCGT
AGGGTACCCGAGAGCGGAGCGTAGAACCCGGTGATGATCACGCTTTCGTTGACGACTTGGAAGCGCTGTGCGAGGA
GGGGCCGGGTGACGGTGGCGCTGGAAGTCCGGTCGAGGTTGTCGCCAGCGGGTTCGGCCAACCTTGCGTGACCCCA
AAGAGGTCAAAGACAGGACCAAAAAAATAAAAAAAAACACAACAGTATTAGGTACCCTGCGCCCCGGGGCCGCCAG
TCCTAATAACACGACGGCGACCGGAAGAACGGCAGTGAGGGCCAGTGCATCTATGGGTGCGGAGCAAGAGAGCGAC
TAACAAAGAGGCCTCATGTTTCATTCCCGGGCGAGCCACACGCGATATCTCGTGCGTCCTTGTTGATTGGTTGAGG
TACAAGCCCTACCCTCCGATTTGCAATCACTTTCGGGACGCGTTTGGGAAACATCTCAGACCGTCACCGAAAAGTT
TAGGGAACTGCGAAAAGACCTGAATCGCCTTCTCTTCCGCTTTTTGCCTTGGCGACCTACAAAAAGGCCACCGCTC
GAGCGCAATCAAACCACCAAACCACAACAAGCGCTACACAGCAGCCAACCATG

Figure 15 ***Acanthamoeba castellanii*** **promoters used for development of an** ***Acanthamoeba*** **specific CRISPR plasmid. A.** U6 spliceosomal RNA promoter. Underlined bases mark the U6 spliceosomal RNA. **B.** Poly-ubiquitin promoter sequence used for driving expression of Cas-9.

In a second step we exchanged the CBh promoter driving Cas9 expression as the CBh promoter is designed for mammalian cell lines (3). We used the polyubiquitin promoter

from *Acanthamoeba* since this promoter was driving protein expression in *Acanthamoeba* in our hands before (S. Deuber, unpublished data) (Figure 15B). The promoter was amplified from *Acanthamoeba* total DNA using the following primers: 5'-CACTCTAGACCGAGCTCTAGCACAAACTC-3' and 5'-CACACCGGTGCATGGTT-GGCTGCTGTGTAG-3' containing an XbaI and AgeI restriction site (underlined bases), respectively. The promoter was introduced into PX459 plasmid using XbaI and AgeI (New England Bioscience). *Acanthamoeba* were not sensitive to puromycin up to 1 mg/ml (Sigma-Aldrich, Buchs, Switzerland) but showed dose-dependent killing in presence of neomycin (G418) (Life Technologies, Carlsbad, CA) (data not shown). In a third step we introduced a cassette for selection of positive *Acanthamoeba* after plasmid transfection. We introduced a kanamycin/ G418 cassette with all necessary promoters and terminators into PX459 using AclI (New England Bioscience) (Figure 14B). After introduction of the *Acanthamoeba* specific promoters and the G418 cassette for selection the plasmid will be used for transfection into *Acanthamoeba* to control Cas9 expression by an anti-FLAG Western blot (Figure 16).

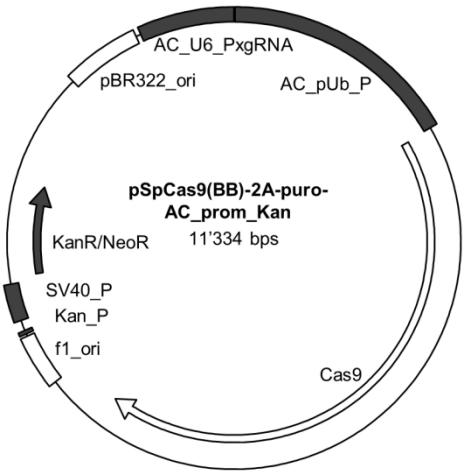


Figure 16 Vector map of an *Acanthamoeba* specific CRISPR plasmid. *Acanthamoeba castellanii* U6 promoter (AC_U6_P), *Acanthamoeba castellanii* polyubiquitin promoter (AC_pUb_P), and the

kanamycin/neomycin resistance cassette (KanR/NeoR) containing all necessary promoters (KanP, SV40P) were introduced into the vector for amoeba specific expression of gRNA, Cas9, and for selection of positive amoeba using neomycin.

Acanthamoeba will be transfected substantially as described in previously (6). Briefly, $5*10^4$ will be seeded in 20 mM Tris pH 8.8, 100 mM KCl, 8 mM $MgSO_4$, 0.4 mM $CaCl_2$, 1 mM $NaHCO_3$ (encystment medium) in 24-wells. In 25 µl encystment medium 1 µg plasmid DNA and 5 µl Lipofectamine2'000 (Life Technologies) will be incubated for 10 min. The mixture will be added to *Acanthamoeba* and incubated for at 30°C 3 h. The transfection mix will then be replaced with growth medium for 24h after which selection will be induced with 50 µg/ml G418. *Acanthamoeba* total cell lysate will be analyzed after expansion for Cas9 expression by anti-FLAG Western blot. As proof-of concept we will knock-out one *Acanthamoeba* gene and one mimivirus gene. In *Acanthamoeba* we chose to knock out actin. Actin fibers are mainly found in lamellipodia and acanthopodia (7). Knock-out of actin leading to a morphologic change in *Acanthamoeba* will be detected by transmission electron microscopy. In mimivirus we chose to knock out L725, a mimivirus fiber protein. Knock-out of L725 leads to a morphologic change in the surface fibers and can be controlled by electron microscopy (8). All gRNAs will be cloned into the plasmid as described in previously (3).

REFERENCES

1. Lander ES. The Heroes of CRISPR. Cell. 2016;164(1-2):18-28. Epub 2016/01/16.

2. Wright AV, Nunez JK, Doudna JA. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell. 2016;164(1-2):29-44. Epub 2016/01/16.

3. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature protocols. 2013;8(11):2281-308. Epub 2013/10/26.

4. Davila Lopez M, Rosenblad MA, Samuelsson T. Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. Nucleic acids research. 2008;36(9):3001-10. Epub 2008/04/09.

5. Clarke M, Lohan AJ, Liu B, Lagkouvardos I, Roy S, Zafar N, et al. Genome of Acanthamoeba castellanii highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. Genome biology. 2013;14(2):R11. Epub 2013/02/05.

6. Peng Z, Omaruddin R, Bateman E. Stable transfection of Acanthamoeba castellanii. Biochimica et biophysica acta. 2005;1743(1-2):93-100. Epub 2005/03/22.

7. Gonzalez-Robles A, Castanon G, Hernandez-Ramirez VI, Salazar-Villatoro L, Gonzalez-Lazaro M, Omana-Molina M, et al. Acanthamoeba castellanii: identification and distribution of actin cytoskeleton. Experimental parasitology. 2008;119(3):411-7. Epub 2008/05/20.

8. Sobhy H, Scola BL, Pagnier I, Raoult D, Colson P. Identification of giant Mimivirus protein functions using RNA interference. Frontiers in microbiology. 2015;6:345. Epub 2015/05/15.

GENERAL DISCUSSION

When mimivirus was first described as virus in 2003 the discovery challenged the classical dogmas of virology. Viruses were always thought to be small, filterable agents (1). This virus, however, was bigger in particle and genome size than various bacteria or archaea, parasitic or free-living. Within the past decade more and more giant viruses from manifold sources infecting various hosts were discovered (2-6). The size and complexity of giant virus genomes raises questions about their host dependence and challenges the separation of viruses in comparison to parasitic cellular organisms. Also, many of the giant virus genes are novel and unexpected for viruses (7). Most protein coding sequences of mimivirus are expressed upon infection of Acanthamoeba indicating a role for these genes in the viral life cycle (8). It seems to be vital for mimivirus to invest energy in the maintenance of its own biosynthetic pathways, among them the glycosylation machinery. The complexity of the mimivirus glycosylation machinery suggests an important role of virus-specific, host-independent glycosylation in the infection of amoeba and the replication of mimivirus. We set out to analyze putative mimivirus glycosyltransferases in order to help understanding of giant virus glycosylation and parallels of giant virus glycosylation to glycosylation in other organisms.

Similarities and differences between mimivirus glycosyltransferases and their eukaryotic or prokaryotic homologs could give valuable insight into the functionality of glycosyltransferases regarding substrate specificity, linkage formation, and pattern recognition on the acceptor substrates. In this work we showed that mimivirus R707 elongated Glc with Glc in both α1,6 and β1,6 and α1,4 configuration while the eukaryotic homolog GYG1 exclusively produces α1,4 linkages between Glc residues (9). GYG1 was crystallized in 2002 and the residues involved in nucleotide sugar donor coordination in the catalytic pocket are known (9). Crystallization of R707 and

comparison of the 3D structure of both proteins might elucidate the mechanism of linkage formation in glycosyltransferases. Mimivirus L230 for example adds Glc residues to polypeptide Hyl while the human homolog GLT25 transfers Gal residues on collagen substrates (10, 11). The exact residues involved in determination of donor specificity of glycosyltransferases are not known to date. Structural analysis of mimivirus glycosyltransferases, their catalytic pocket and steric changes upon donor and acceptor substrate binding together with a thorough sequence analysis could answer questions regarding the mechanism of glycosyltransferase catalysis in general.

LIMITATIONS IN GLYCOSYLTRANSFERASE ANALYSIS

Considering that 1-2% of all genes are encoding glycosyltransferases studies of glycans lag behind those of other major classes of molecules, like DNA or proteins (12). Glycans are highly complex, their structure can neither predicted from a template nor does universal code for glycans exist. A certain glycosylation site on a protein may be glycosylated or not, and different copies of one protein may vary in glycan structure on one specific glycosylation site (13). This hampers the study of glycans. In addition, possibilities to analyze the complexity of glycans were restricted by a shortage of efficient methods. With the development of advanced technologies with powerful computational tools the study of glycans has become more feasible (14). Limitations in analysis of glycosyltransferases are anyhow still prevailing to date. While the history of protein databases reaches back to the 1970s the CAZY database went online in 1998 (15-19). On CAZY, glycosyltransferases sharing at least 100 amino acids in two different stretches with members of one glycosyltransferase family but not to other families this glycosyltransferase are assigned to that specific family with the same predicted glycosyltransferase fold (20). More than 95% of the annotated glycosyltransferases, however, are uncharacterized (12). The putative mimivirus glycosyltransferases, for example, were grouped within family 1 (3.5% characterized), family 2 (0.5%

characterized), family 8 (1.0% characterized), family 10 (7.5% characterized), and family 32 (1.0% characterized). Within some families the range of glycosyltransferase activity is narrow, while members of another family have a broad range of activities using different nucleotide sugar donors, acceptor substrate, form different linkages, and originate from different organisms. Glycosyltransferase family prediction can only give hints towards the functionality of a certain enzyme. Sequence analysis by BLAST can predict certain domains and functionalities. These results, however, are not always revealing. The same holds true for protein analysis with tools taking 3D folds into account, like Phyre[2]. In case of the analyzed putative mimivirus glycosyltransferases family and domain prediction was helpful in 2 out of 6 cases. Mimivirus R707 was predicted to be a GYG1 homolog and indeed R707 was active as elongating glucosyltransferase. Linkage formation, however, was different in R707 catalysis. Mimivirus L373 showed strongest homologies to fungal mannosyltransferases and an N-acetylglucosaminyltransferase from S. pneumonia. L373 hydrolyzed UDP-GlcNAc, but not GDP-Man. Mimivirus R654 was predicted to be a fucosyltransferase. Fuc is only present in traces in mimivirus. We could not detect any GDP-Fuc hydrolysis or fucosyltransferase activity. All other investigated putative mimivirus glycosyltransferases were grouped with uncharacterized proteins and families with various activities. For those candidates, the bioinformatics tools available to date failed to help unraveling the functionality of the glycosyltransferases.

Another challenge in determination of mimiviral glycosyltransferase activity is the exact acceptor substrate specificity of most glycosyltransferases like it is seen for example in the human blood group glycosyltransferases (21). To identify a mimivirus core glycosyltransferase a library of all mimivirus glycoproteins would be needed. To date such a list does not exist and in addition it would be cumbersome to clone and recombinantly express all these acceptor proteins. We tried various mimivirus peptides with known glycosylation sites as acceptor substrates. As glycosylation in mimivirus

might occur on fully folded proteins a peptide alone might not be recognized by the viral glycosyltransferase. Even with a number of glycan structures known it is hard to get hands on these complex glycan structures as acceptor substrates for elongating glycosyltransferases. With nothing known to date about mimivirus N-glycans the hunt for glycosyltransferase activity is even more complicated.

In addition, many glycosyltransferases have both N- and O-glycosylation sites themselves (22). The consensus glycosylation sequences and glycan structures of eukaryotic and prokaryotic glycosylation differ from those in giant viruses (23, 24). Recombinant expression of putative mimivirus glycosyltransferases in a eukaryotic or prokaryotic system might hinder proper glycosylation of mimivirus glycosyltransferases which might be needed for their activity.

IN VIVO STUDIES OF MIMIVIRUS GLYCOSYLTRANSFERASES

Given the fact that mimivirus is one of very few viruses encoding an own glycosylation machinery it would be interesting to study the role of mimivirus glycosyltransferases *in vivo* to determine the role of mimivirus glycosylation in infection and replication. Depletion of one or a systematic depletion of multiple putative glycosyltransferases would allow in depth studying of the role of mimivirus glycosylation in viral infection and replication. One could study the changes in surface glycosylation and infection and replication rate. A study using RNA interference in mimivirus was published in 2015 (25). RNA interference, however, is only transient and does not allow virus expansion and propagation in large quantities. Genetic engineering of large DNA viruses, like herpesviruses and poxviruses, was done in the past using cosmids and bacterial artificial chromosomes (26). These techniques, however, are laborious and time-consuming, and the mimivirus genome is almost 5-10 folds bigger than the genome of herpesviruses. The RNA-guided Cas9 nuclease from microbial CRISPR adaptive immune system is used for genome engineering in a broad range of species (27-29). The

CRISPR-Cas9 is successfully used for herpesvirus generating specific mutated viruses within two weeks (30, 31). Adaptation of a CRISPR plasmid for Cas9 and gRNA expression in *Acanthamoeba* would allow us to targeting any mimivirus gene for knock-out by cloning of a mimivirus specific gRNA into the plasmid. Functionality of the plasmid we generated for *Acanthamoeba* specific expression of Cas9 and the gRNA needs to be validated *in vivo*. A mimivirus specific CRISPR-Cas9 would open the door to studies of not only mimivirus glycosyltransferases but any kind of gene or whole groups of genes giving valuable insights about the functionality of mimivirus proteins.

MIMIVIRUS GLYCOSYLTRANSFERASES IN GLYCOENGINEERING

It is accepted by now that glycan structures play crucial biological and physiological roles. Glycosylation became a major concern in biopharmaceutical manufacturing, production of therapeutically relevant proteins and antibodies, and other bioactive molecules (13, 32-34). Glycoengineering can enhance *in vivo* activity and alter pharmacokinetic properties of therapeutic proteins (35). The production of certain, defined glycans, however, is still challenging. Microorganisms can be used as platforms for production of large amounts of therapeutic proteins. Using microbes is cheaper, more rapid and stable in comparison to protein production in mammalian cell lines. The produced proteins often show host glycan patterns which may differ from the desirable glycan patterns making glycoengineering necessary. Microbes can be used as cell factories for the production of proteins and the glycan structures can afterwards be added *in vitro* (36). Mimivirus glycosyltransferases could be produced and purified in reasonable quantities *in vitro*. The analyzed glycosyltransferases did not have transmembrane domains and it is therefore likely that they do not need to be embedded in an organelle membrane like ER or Golgi for proper functioning. This makes them good candidates for use in *in vitro* glycosylation. Mimivirus glycosyltransferases would be especially useful for core O-glycosylation and elongation of both O- and N-glycans.

As the sequon in giant virus N-glycosylation differs drastically from eukaryotic and prokaryotic sequons, putative mimivirus core N-glycosylation enzymes might not be useful for core glycosylation of eukaryotic target proteins (37, 38).

MIMIVIRUS GLYCOSYLATION IN IMMUNOGENICITY

Giant viruses are found in both fresh water and marine environment, they were isolated from soil and permafrost (1, 6, 7, 39). In case of mimivirus humans might come in contact with the virus by ingesting infected water or by inhaling virus particles as aerosols. In both cases mimivirus is probably taken up by macrophages which can trigger presentation of mimivirus peptides to the immune system (40). Many mimivirus proteins have eukaryotic origin which enhances the chances of cross-reactivity with endogenous proteins upon activation of the immune system (41). There are various examples of molecular mimicry caused by pathogen infection of humans. Autoimmunity is a common feature of hepatitis B and C virus infection inducing the production of autoantibodies which cause a broad spectrum of autoimmune diseases like cryoglobulinaemia, autoimmune thyroiditis, rheumatoid arthritis, and more (42). *Streptococcus* infection is associated with rheumatic fever induced by molecular mimicry of streptococcal group A M proteins and myosin (43, 44). These cases of molecular mimicry are protein or peptide related. A prominent example of carbohydrate mimicry is seen in *Campylobacter jejuni* infection. Carbohydrate mimicry between human ganglioside GM1 and a lipo-oligosaccharide of *C. jejuni*

induces Guillain-Barré syndrome in patients (45, 46). A study in our group showed that mimivirus infection induces rheumatoid arthritis in mice. Immunization of mice with mimivirus total protein extract leads to generation of auto-reactive anti-collagen antibodies accompanied by rheumatoid arthritis syndromes. Interestingly, human sera of arthritis patients react to mimivirus collagen in a significantly higher level than sera from a healthy control group indicating a production of auto-antibodies upon exposure

to mimivirus(47). Given the fact that some mimivirus glycosyltransferases have eukaryotic homologs and most glycan structures are found on the surface of mimivirus particles it is possible that other mimivirus surface glycans induce autoimmune reactions making it particularly interesting to study und understand the synthesis and structures of the glycan shell of mimivirus.

Mimivirus is a peculiar organism with manifold special features never seen in a virus before. We were able to gather further evidence for an independent glycosylation machinery in mimivirus. Identifying the functionality of R707 depicts an exception to the common rule of the functionality of a glycosyltransferase. R707 is not only flexible towards the linkage formed, which is known for a few other glycosyltransferases, R707 is also capable of forming both α and β linkages, a feature never described for glycosyltransferases before. This finding might help understanding the general mechanism of glycosidic bond formation and underlines the importance of basic research in curious organisms like giant viruses.

REFERENCES

1. Rivers TM. Viruses and Koch's Postulates. Journal of bacteriology. 1937;33(1):1-12. Epub 1937/01/01.

2. Abergel C, Legendre M, Claverie JM. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. FEMS microbiology reviews. 2015;39(6):779-96. Epub 2015/09/24.

3. Campos RK, Boratto PV, Assis FL, Aguiar ER, Silva LC, Albarnaz JD, et al. Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. Virology journal. 2014;11:95. Epub 2014/06/03.

4. Colson P, Gimenez G, Boyer M, Fournous G, Raoult D. The giant Cafeteria roenbergensis virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. PloS one. 2011;6(4):e18935. Epub 2011/05/12.

5. Colson P, Yutin N, Shabalina SA, Robert C, Fournous G, La Scola B, et al. Viruses with more than 1,000 genes: Mamavirus, a new Acanthamoeba polyphaga mimivirus strain, and reannotation of Mimivirus genes. Genome biology and evolution. 2011;3:737-42. Epub 2011/06/28.

6. Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, et al. Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome biology and evolution. 2012;4(12):1324-30. Epub 2012/12/12.

7. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. Science. 2004;306(5700):1344-50. Epub 2004/10/16.

8. Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D, et al. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. Genome research. 2010;20(5):664-74. Epub 2010/04/03.

9. Gibbons BJ, Roach PJ, Hurley TD. Crystal structure of the autocatalytic initiator of glycogen biosynthesis, glycogenin. Journal of molecular biology. 2002;319(2):463-77. Epub 2002/06/08.

10. Luther KB, Hulsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennet T. Mimivirus collagen is modified by a bifunctional lysyl hydroxylase and glycosyltransferase enzyme. The Journal of biological chemistry. 2011. Epub 2011/11/03.

11. Schegg B, Hulsmeier AJ, Rutschmann C, Maag C, Hennet T. Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. Molecular and cellular biology. 2009;29(4):943-52. Epub 2008/12/17.

12. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. Annual review of biochemistry. 2008;77:521-55. Epub 2008/06/04.

13. Spahn PN, Lewis NE. Systems glycobiology for glycoengineering. Current opinion in biotechnology. 2014;30:218-24. Epub 2014/09/10.

14. Varki A, Sharon N. Historical Background and Overview. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

15. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic acids research. 2009;37(Database issue):D233-8. Epub 2008/10/08.

16. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic acids research. 2014;42(Database issue):D490-5. Epub 2013/11/26.

17. Campbell JA, Davies GJ, Bulone V, Henrissat B. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. The Biochemical journal. 1997;326 (Pt 3):929-39. Epub 1997/10/23.

18. Coutinho PM, Deleury E, Davies GJ, Henrissat B. An evolving hierarchical family classification for glycosyltransferases. Journal of molecular biology. 2003;328(2):307-17. Epub 2003/04/15.

19. Berman HM. The Protein Data Bank: a historical perspective. Acta crystallographica Section A, Foundations of crystallography. 2008;64(Pt 1):88-95. Epub 2007/12/25.

20. Brockhausen I. Crossroads between Bacterial and Mammalian Glycosyltransferases. Frontiers in immunology. 2014;5:492. Epub 2014/11/05.

21. Rini J, Esko J, Varki A. Glycosyltransferases and Glycan-processing Enzymes. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

22. Varki A, Esko JD, Colley KJ. Cellular Organization of Glycosylation. In: Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, et al., editors. Essentials of Glycobiology. 2nd ed. Cold Spring Harbor (NY)2009.

23. De Castro C, Molinaro A, Piacente F, Gurnon JR, Sturiale L, Palmigiano A, et al. Structure of N-linked oligosaccharides attached to chlorovirus PBCV-1 major capsid protein reveals unusual class of complex N-glycans. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(34):13956-60. Epub 2013/08/07.

24. Hulsmeier AJ, Hennet T. O-Linked glycosylation in Acanthamoeba polyphaga mimivirus. Glycobiology. 2014;24(8):703-14. Epub 2014/05/06.

25. Sobhy H, Scola BL, Pagnier I, Raoult D, Colson P. Identification of giant Mimivirus protein functions using RNA interference. Frontiers in microbiology. 2015;6:345. Epub 2015/05/15.

26. Flint SJE, L.W.; Racaniello, V.R.; Skalka, A.M. Principles of Virology. 3rd ed. Washington, D.C.: ASM Press; 2009. 569 p.

27. Lander ES. The Heroes of CRISPR. Cell. 2016;164(1-2):18-28. Epub 2016/01/16.

28. Wright AV, Nunez JK, Doudna JA. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell. 2016;164(1-2):29-44. Epub 2016/01/16.

29. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature protocols. 2013;8(11):2281-308. Epub 2013/10/26.

30. Bi Y, Sun L, Gao D, Ding C, Li Z, Li Y, et al. High-efficiency targeted editing of large viral genomes by RNA-guided nucleases. PLoS Pathog. 2014;10(5):e1004090. Epub 2014/05/03.

31. Suenaga T, Kohyama M, Hirayasu K, Arase H. Engineering large viral DNA genomes using the CRISPR-Cas9 system. Microbiology and immunology. 2014;58(9):513-22. Epub 2014/07/22.

32. Strasser R, Altmann F, Steinkellner H. Controlled glycosylation of plant-produced recombinant proteins. Current opinion in biotechnology. 2014;30:95-100. Epub 2014/07/08.

33. Xiao J, Muzashvili TS, Georgiev MI. Advances in the biotechnological glycosylation of valuable flavonoids. Biotechnology advances. 2014;32(6):1145-56. Epub 2014/05/02.

34. Herter S, Birk MC, Klein C, Gerdes C, Umana P, Bacac M. Glycoengineering of therapeutic antibodies enhances monocyte/macrophage-mediated phagocytosis and cytotoxicity. J Immunol. 2014;192(5):2252-60. Epub 2014/02/04.

35. Sinclair AM, Elliott S. Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. Journal of pharmaceutical sciences. 2005;94(8):1626-35. Epub 2005/06/17.

36. Anyaogu DC, Mortensen UH. Manipulating the glycosylation pathway in bacterial and lower eukaryotes for production of therapeutic proteins. Current opinion in biotechnology. 2015;36:122-8. Epub 2015/09/05.

37. Van Etten JL, Gurnon JR, Yanai-Balser GM, Dunigan DD, Graves MV. Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the endoplasmic reticulum and Golgi. Biochimica et biophysica acta. 2010;1800(2):152-9. Epub 2009/08/06.

38. Zhang Y, Xiang Y, Van Etten JL, Rossmann MG. Structure and function of a chlorella virus-encoded glycosyltransferase. Structure. 2007;15(9):1031-9. Epub 2007/09/14.

39. Philippe N, Legendre M, Doutre G, Coute Y, Poirot O, Lescot M, et al. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science. 2013;341(6143):281-6. Epub 2013/07/23.

40. Eric Ghigo JK, Oham Lien, Lucas Pelkmans, Christian Capo, Jean-Louis Mege, Didier Raoult. Ameobal Pathogen Mimivirus Infects Macrophages through Phagocytosis. PloS Pathogens. 2008.

41. Yutin N, Wolf YI, Koonin EV. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology. 2014. Epub 2014/07/22.

42. Bogdanos DP, Mieli-Vergani G, Vergani D. Virus, liver and autoimmunity. Digestive and liver disease : official journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver. 2000;32(5):440-6. Epub 2000/10/13.

43. Cunningham MW, McCormack JM, Fenderson PG, Ho MK, Beachey EH, Dale JB. Human and murine antibodies cross-reactive with streptococcal M protein and myosin recognize the sequence GLN-LYS-SER-LYS-GLN in M protein. J Immunol. 1989;143(8):2677-83. Epub 1989/10/15.

44. Stollerman GH. Rheumatic fever. Lancet. 1997;349(9056):935-42. Epub 1997/03/29.

45. Ang CW, Jacobs BC, Laman JD. The Guillain-Barre syndrome: a true case of molecular mimicry. Trends in immunology. 2004;25(2):61-6. Epub 2004/04/23.

46. Yuki N, Susuki K, Koga M, Nishimoto Y, Odaka M, Hirata K, et al. Carbohydrate mimicry between human ganglioside GM1 and Campylobacter jejuni lipooligosaccharide causes Guillain-Barre syndrome. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(31):11404-9. Epub 2004/07/28.

47. Shah N, Hulsmeier AJ, Hochhold N, Neidhart M, Gay S, Hennet T. Exposure to mimivirus collagen promotes arthritis. Journal of virology. 2014;88(2):838-45. Epub 2013/11/01.

ACKNOWLEDGMENTS

I would like to thank my supervisor Prof. Thierry Hennet for giving me the opportunity to be a PhD student in his research group and for always having an open door.

I would like to thank the members my thesis committee Prof. Lubor Borsig, Prof. Michela Tonetti, and Prof. Urs Greber for the scientific input and helpful suggestions during my committee meetings.

Thanks to Simon Jurt his help with all NMR related matters and to Dr. Cinzia Bernardi for her help in cloning.

During my PhD I had the opportunity to work together with great minds and learn a lot from them. Special thanks to Dr. Andreas Hülsmeier, Dr. Kelvin Luther, Dr. Jürg Cabalzar and Christoph Rutschmann for teaching me so much, always answering my questions and for all the scientific and non-scientific discussions we had.

It is a great, when colleagues become friends. Thank you Adrienne for all experiences we shared. Thanks to the running, body combat and work-out team, Nik, Michi, Nina, Sacha, Katharina, Jesus. Thanks to current and former group members for creating a nice working environment and the fun we had together: Eddie, Marek, Stephan, Luca, Yiling, Thomas. I would also like to thank my current and former office members, Darya, Cristina, Jesus, Marko, and Irina. Thank you for making me feel at home in L14. It Special thanks to Giovi not only for administrative help but for always listening. Special thanks to my former colleague and friend Aurelia for always understanding me.

In the end I would like to thank my friends and family, without whom I would not have been able to finish my thesis. All of you helped and supported me in many different ways. Thank you Anki and Oskar for always being there for me, I love you a lot. Thank you Jürg, the best is yet to come.

CURRICULUM VITAE

Education

2011-2016	PhD in biology, University of Zurich, Institute of Physiology, Group of Prof. Hennet: Characterization of putative glycosyltransferases in giant *Acanthamoeba polyphaga* mimivirus
2009-2011	Master of Science, major in cell biology, ETH Zurich, University Hospital Zurich, Group of Dr. Beer: The role of reactive oxygen species in inflammasome activation in human keratinocytes
2005-2009	Bachelor of Science in chemical biology, ETH Zurich
2002-2005	Matura, major in biology/chemistry, Kantonsschule ImLee Winterthur

Publications

Rommel AJ., Hülsmeier AJ., Jurt S., Hennet T., Giant mimivirus R707 encodes a glycogenin paralog polymerizing glucose through alpha and beta glycosidic linkages. Biochemical Journal, in press.

Rommel AJ., Hennet T., *Acanthamoeba polyphaga* mimivirus encodes a UDP-GlcNAc hydrolyzing enzyme with glycosyltransferase characteristics. Manuscript in preparation.

Funding

2012-2013	Candoc Research Credit, University of Zurich. 50'000 CHF

Conference activities

2015	Gordon Research Conference Glycobiology, Lucca, Italy Poster: Characterization of putative glycosyltransferases in giant *Acanthamoeba polyphaga* Mimivirus
2015	Gordon Research Seminar Glycobiology, Lucca, Italy Talk: Characterization of putative glycosyltransferases in giant *Acanthamoeba polyphaga* Mimivirus
2013	LS2 Annual Meeting, Zurich, Switzerland Poster: Characterization of Mimivirus glycosyltransferases